The Theory of the Construction of Tables of Mortality

AND OF

Similar Statistical Tables in use by the Actuary.

A COURSE OF LECTURES

BY

GEORGE FRANCIS HARDY,

FELLOW OF THE INSTITUTE OF ACTUARIES.

DELIVERED AT THE

Institute of Actuaries, Staple Inn Hall,

During the Session, 1904-5.

Published for the Institute of Actuaries by

CHARLES AND EDWIN LAYTON,

56, FARRINGDON STREET, LONDON, E.C.

1909.

PREFATORY NOTE.

To each set of Lectures delivered before the Institute of Actuaries, when published in book form, there has generally been prefixed a short preface, or introduction, written by the President of the Institute then in office. This course, admirable in itself, cannot well be followed on the present occasion, having regard to the fact that MR. HARDY has, in the interval between the delivery of the Lectures and their publication, himself been elected to the Presidential chair. It has therefore devolved upon us, as Honorary Secretaries of the Institute, to insert this foreword in explanation of a seeming omission, and to express therein the confidence of the Council that the Lectures will be found to be of the greatest interest and value to the profession, which already owes so deep a debt of gratitude to their author.

J. E. F.

W. P. P.

PREFACE.

THE object of the following Lectures was to deal with the theoretical considerations that should govern the selection and treatment of such statistics as form the basis of the various tables of mortality, sickness, secession, marriage, superannuation, etc., which are of use to the Actuary. It should be noted that in nearly all cases where mortality tables are specially referred to what is said may be extended to other types of statistics, though, to avoid repetition, that is not always pointed out.

Some apology is required for the long delay in the publication of the Lectures. It was intended subsequently to their delivery, to expand them into something like a complete treatment of the subject (from the theoretical point of view), and to add a sufficient series of examples to illustrate the various points of theory. Unfortunately I have not found time to carry out this intention, but as regards that part of the subject dealing with the use of the Pearsonian Types of Frequency Curves in Statistics this has been rendered unnecessary by the appearance of Mr. ELDERTON's admirable book upon "Frequency Curves and Correlation", published by the Institute of Actuaries in 1906.

A few additions have, however, been made to the Lectures as originally delivered, and where these appeared to interfere with the continuity of the text they have been relegated to notes placed at the end of the Lectures.

I have very specially to thank Mr. G. J. LIDSTONE, F.I.A., for several valuable suggestions, in particular for the contribution of Notes, and for assistance in preparing the lectures for the Printers; and also Dr. JAMES BUCHANAN, M.A., F.I.A., F.F.A., for having kindly revised the proofs and checked the algebra and numerical work.

G. F. H.

The Theory of the Construction of Tables of Mortality

AND OF

Similar Statistical Tables in use by the Actuary.

BY

G. F. HARDY, F.I.A.

FIRST LECTURE.

WHEN the Council asked me to deliver a series of lectures upon some subject connected with Part III of the Institute Examination I selected the construction of mortality and similar statistical tables, mainly because it seemed to me to lie at the basis of our work. Actuarial science, in the modern sense of the term, had its origin in the collection of statistics (however rough and inaccurate these may have been), and their use for the purpose of calculating life contingencies; and although the Actuary has now to take account of a wider range of subjects than formerly, the collection and analysis of past experience and the employment of the results of such analysis to forecast the future is still his most important function.

The title of the lectures is somewhat wider and more ambitious than the contents may be found to warrant. To justify it fully would involve dealing with many questions of detail relating to the collection and tabulation of data, such, for example, as the various methods for computing the numbers exposed to risk in a mortality experience, &c., which have been many times discussed in the volumes of the *Journal of the Institute of Actuaries* and many of which are exhaustively dealt with by Mr. Ackland in the recently published "Account of Principles and Methods." It is evident that to deal with the subject in such detail, would outrun the limits of the six lectures which I have undertaken to deliver. I propose, therefore, to confine myself mainly to

a consideration of the general principles involved in the collection of statistical data, and in the construction from such data of tables, of which the Mortality Table is the best known and the most important, embodying the results in the form required by the Actuary, and, at the same time, to give such examples of the application of these principles as may be necessary to illustrate the subject.

In this opening lecture in particular, I shall ask your indulgence if occasionally my remarks appear to be of an elementary character, as I think it desirable that we should be perfectly clear as to first principles before going on to more detailed consideration of the subject.

Statistical tables, in one form or another, are familiar to all of us. At the basis of all such tables, and, indeed, of the whole science of statistics, lies one of the most fundamental facts in nature, namely, that all phenomena of which we have any knowledge fall into certain classes, groups or series, and cluster round certain types. But for this fact we should be unable to classify our knowledge, indeed, should never have acquired any to classify. Speaking broadly, then, every object and every event that comes within our observation is one of a group or class of similar but not identical objects or events, which, as a class, is marked off by certain special features from every other class, although the dividing line may not always be sharply drawn. These groups or classes are not arbitrary, but are inherent in the nature of things, although it is true that the particular groups which we employ in classifying our knowledge are chosen with a view to our own convenience and to the limitations of our minds.

From a consideration of a class of objects as a whole, we get a conception of an average, or type,* to which each individual in the class more or less conforms, but from which, notwithstanding, every individual also diverges. Such divergencies or variations of individuals from the average type may be discontinuous, themselves running into types, or they may be continuous. Among the individuals forming together the type mankind, are divergencies such as those due to sex, race, nationality, birthplace, occupation, civil condition, &c., discontinuous variations producing sub-groups, the boundaries of which overlap and interlace, each

* The type of the class should preferably be considered as represented by the "mode" or case of most frequent occurrence rather than by the "average" or "mean", but this point is not here of importance.

of these smaller groups again being capable of endless subdivision. These divergencies can be dealt with statistically only by counting the members of the various sub-groups.

On the other hand, there are divergencies, which we may term continuous, such as those due to differences of age, height, weight, income, &c., &c., differing from the former class in that they do not involve the separation of the main group into sub-groups, but relate to qualities, possessed by each member of the group in varying degree, capable of measurement and numerical statement, and involving the idea in each instance of an average. Thus we can speak of the average age, height, or income of a group of persons, not of their average occupation or nationality, although we may speak of the average constitution of the group in respect of these latter qualities.

A statistical table deals with some natural group of objects or events and is a numerical statement of the manner in which the members of the particular group differ *inter se* in respect of some special character or characters. If dealing with discontinuous variations, as for example a table showing the occupations of a group of persons, it will exhibit, implicitly or explicitly, the ratio of the magnitude of each sub-group to the whole, at a given moment or moments or on an average of a given period; or it may take the form of a statement of the extent to which variations in one respect are affected by variations in another, as, for example, a table showing the proportion of the sexes in different nationalities. If dealing with continuous variations, it will either represent a series of measurements of some quality common to members of the group, showing its average value for the group, and the manner in which individual values are grouped round such average, or it may represent, numerically, the manner in which deviations from the average in respect of some one quality A are correlated with the deviations in respect of some other quality B.

It is mainly with the class of statistical table dealing with continuous variations that the Actuary has to deal; variations in the ages of lives under observation, their ages, or the periods elapsed since entry, at death, withdrawal, marriage, superannuation, &c. In such tables the grouping of individual measures round the average will, in general, but not always, be found to follow, approximately, certain well-defined laws. Taking first the tables dealing with a single variable, the

following may be considered as an example. It is a statement of the heights of 2,192 school children, and is abridged from that given in a paper by Prof. Karl Pearson.

TABLE I.

Showing heights of 2,192 School Children, aged 12 years.

Heights in Centimetres	No. of Children Observed	Computed Nos. by Curve $\kappa e^{-x^2/c^2}$	Computed—Observed	
			+	−
(1)	(2)	(3)	(4)	(5)
139–140	1	...	...	1
135–138	6	3	...	3
131–134	31	25	...	6
127–130	107	119	12	...
123–126	321	338	17	...
119–122	585	577	...	8
115–118	618	596	...	22
111–114	359	365	6	...
107–110	126	135	9	...
103–106	35	30	...	5
99–102	3	4	1	...
Total	2,192	2,192	45	45

NOTE.—In the formula (col. 3) x represents the deviation in centimetres from the average; $c=7{\cdot}76$ and κ has such a value $\frac{2192}{c\sqrt{\pi}}$ as to make the area of the graduated curve equal to the ungraduated; that is, to make the totals of columns (2) and (3) equal.

If we consider the progression of the numbers in column (2), we shall see that they form a roughly symmetrical series, being largest in the neighbourhood of the average height and diminishing gradually on either side. It will be seen that the average height is about $118\frac{1}{4}^{c}$, the number exceeding this height being approximately equal to the number falling short of it. In order to bring out the approximate law of the series, I have inserted in column (3) the computed numbers on the assumption that the frequency of a deviation of $\pm x$ centimetres from the average is represented by the function $\kappa e^{-x^2/c^2}$, where c has the value 7·76 and κ the value $\frac{2192}{c\sqrt{\pi}}$. The expression $\kappa e^{-x^2/c^2}$, represents what is usually termed the curve of "facility of error", or the "normal" curve of frequency. It will be seen that while the figures in column (2), are as we should expect them to be with such limited data, somewhat irregular, they conform on the whole fairly closely to the normal curve.

The "normal curve" was first used to represent the distribution as to magnitude of errors of observation in physical measurements. It must not be regarded as representing a law of Nature, but rather an extremely convenient and often very close approximation to observation; experience proving that in many cases errors of observation and the deviations of individuals from the mean of a class do follow very closely the law referred to. The formula is therefore empirical and not to be established by *a priori* reasoning; at the same time we may, perhaps, see a logical basis in the following consideration. We may suppose that, in any individual measurement, the deviation from the mean of the class (as the difference in the height of any individual among the 2,192 in Table I from the average height of the whole group) is the result of an infinity of minute causes as to whose nature we are in ignorance, any one of which may produce a minute positive or negative deviation from the average. These minute superimposed deviations being indefinitely small and indefinitely numerous, we may without loss of generality assume them of equal magnitude. It is then clear that the magnitude and sign of the total resulting deviation in any given case will depend upon the extent to which the number of these minute positive deviations exceed the negative, or *vice versa*.

If the number of possible causes of deviation is $2n$, and if the extent of each indefinitely small deviation is k (n being indefinitely large, but $k\sqrt{n}$ finite), then the probability or "frequency" of a total deviation lying between x and $x+k$ will depend on our having $\left(n+\frac{x}{2k}\right)$ positive values of k and $\left(n-\frac{x}{2k}\right)$ negative values. The probability of this occurring will be represented by the appropriate term in the expansion of the binomial $(\frac{1}{2}+\frac{1}{2})^{2n}$ or

$$\frac{\underline{|2n}}{\underline{\left|n+\frac{x}{2k}\right.}\;\underline{\left|n-\frac{x}{2k}\right.}}\left(\tfrac{1}{2}\right)^{2n}.$$

It may easily be shown that this expression, n being indefinitely great, takes the form

$$\frac{1}{\sqrt{\pi n}}\,e^{\frac{-\left(\frac{x}{2k}\right)^2}{n}},\ i.e.\ (\text{Constant})\times e^{-\frac{x^2}{c^2}}$$

i.e., of the curve of the "facility of error." I do not propose to discuss at any length the properties of this particular curve,‡ but you will notice that the curve being symmetrical with respect to positive and negative values of x, it assumes that positive and negative deviations of a given magnitude are equally frequent, the average magnitude of such deviations being small or large as c is small or large. The maximum ordinate corresponds to the value of $x=0$, which is the average value of x; it therefore passes through the centre of gravity of the area enclosed by the curve and the axis of x, and also divides that area into two equal parts. It assumes that indefinitely large deviations are possible, hence it cannot be rigidly exact, because when dealing with physical measurements of any kind, indefinitely large errors are not possible. This is not a practical objection to the use of the formula, however, as the probability thereunder of deviations of many times the average value is extremely small.

The following table, showing the number of entrants in various aged groups in the O[M] Experience, exhibits a quite different distribution of the deviations from the average:

TABLE II.

Number of entrants in quinary age groups O[M] data.

Central Age of Group x	Actual Entrants in Group*	Computed No. by Formula†	Computed—Actual +	Computed—Actual −
(1)	(2)	(3)	(4)	(5)
20	431	436	5	...
25	1,273	1,305	32	...
30	1,526	1,473	...	53
35	1,269	1,265	...	4
40	914	930	16	...
45	591	604	13	...
50	354	349	...	5
55	182	178	...	4
60	83	79	...	4
65	26	29	3	...
70	7	8	1	...
75	1	1	...	...
Totals	6,657	6,657	70	70

* Omitting hundreds.

† Formula representing number of entrants at given age $x=\kappa(x-18{\cdot}59)^{1{\cdot}041}(88{\cdot}48-x)^{6{\cdot}004}$; where $\log \kappa = -9{\cdot}2360$.

‡ The student may consult Woolhouse's paper on "The Philosophy of Statistics" (*J.I.A.*, vol. xvii, p. 37), or an exhaustive analysis of the properties of the curve by Mr. Sheppard (*Phil. Trans.*, vol. 192, p. 101); *See* also "Bowley's Elements of Statistics", Part II, Sec. II.

Here the numbers also exhibit a well-marked law governing the deviations from the mean, but this law is no longer the same as that shown by the "normal" curve of frequency. The maximum ordinate does not coincide either with the average age or with the central age of the series; while the number of cases exceeding the average age no longer equals the number falling short of it. In other words, the curve is non-symmetrical or skew. It follows very approximately, however, a certain law, as will be seen by comparing the numbers in column (2) with those in column (3), which represent the computed numbers according to the formula stated.

Having regard to the fact that the numbers in column (2) represent 100's and not units, the differences between the actual and computed numbers are somewhat outside the probable errors of observation. There are, that is to say, "systematic" differences between the two curves. These systematic differences are generally to be expected in dealing with age statistics. It will be seen that they are not incompatible with a close agreement in the general features of the two curves, but they serve as a warning that, in statistics of this nature, formulæ representing the distribution of deviations from the mean must be regarded as approximations only.

If we consider the curves exhibited in Tables I and II we see that the general character of such curves is determined by a few salient features:

1. The position of the maximum ordinate; that is, the value of the variable having maximum frequency. This value is termed the *mode*.
2. The average or *mean* value of the variable, being the arithmetical mean of all individual values. In a symmetrical curve this coincides with the "mode."
3. The *average deviation* from the mean, corresponding to the closeness with which the individual measures are grouped round their mean value. There is a certain convenience, for analytical reasons, in adopting as our standard in this respect either the mean of the squares of the individual deviations, or the square root of this quantity. The latter is termed the *standard deviation*. We may represent the average of the squares of the deviations, or the

"mean square" deviation by the symbol μ_2, when the *standard deviation* becomes $\sqrt{\mu_2}$.

4. The equality or otherwise of the positive and negative deviations from the mean; that is, the *symmetry* or *skewness* of the curve. The sum of the first powers of the deviations is, of course, always zero. If the curve is symmetrical, the sum of any odd power of the deviations must be zero, but not otherwise. As we have employed the square root of the average square of the deviations as a measure of the diffuseness or spread of the curve, termed the "standard deviation", so we may take the ratio of the cube root of the average cube deviation to the "standard deviation" as the standard of "skewness." If we represent the average cube deviation by the symbol μ_3, the *skewness* of the curve may then be measured by $\frac{\sqrt[3]{\mu_3}}{\sqrt{\mu_2}}$.

The skewness is sometimes taken as the difference between the "mean" and the "mode", divided by the standard deviation.

The sums of the successive powers of the deviations of the variable from the mean, the area of curve being taken as unity, are termed the *moments* of the curve.

These observed laws of the variation of measurements from their mean are very general, and are usually, though not invariably, associated with what is termed "homogeneous" data. The distinction between "homogeneous" and "heterogeneous" data is of considerable importance, although not very easy to define. We may perhaps define a homogeneous group as one in which the continuous variations are from a single type only, and are unaffected by any discontinuous variations in the group if these exist. These conditions will hardly ever prevail, but a group may be considered for practical purposes as homogeneous if the variations in the particular quality dealt with are not materially affected by any discontinous variations existing in the group. If, however, the group can be split up into two or three distinct series differing markedly in certain qualities, and these differences are found, or may reasonably be supposed, to affect the character under examination, then the series is "heterogeneous."

Take, for example, the class representing assured lives of

a given age, but of varying duration of assurance, and assume we are investigating the rate of mortality of the class. If it is found on examination that the duration of assurance materially affects the rate of mortality, then the data treated as a whole is heterogeneous. If it is found, however, that the duration of assurance after reaching a certain point has no such influence, or an influence that is insignificant, then the data from this point and in this respect may be treated as homogeneous. The same considerations apply to distinctions in class of assurance, amount of policy, occupation, &c.

The laws which appear to govern deviation from the average in homogeneous data are, in general, so uniform in action that a departure therefrom will frequently indicate that data which might be supposed to be homogeneous are not so. An interesting illustration of this may be seen in the case of the Male Annuitants in the New Offices' Annuity Experience. Consider the following table showing the number of entrants for various groups of ages:—

TABLE III.

MALE ANNUITANTS O^{am} DATA.

Number of entrants at various ages, 1863-1893.

Ages at Entry x	Entrants 1863-1893	Computed Numbers $n_x = \kappa e^{-\left(\frac{x-65}{c}\right)^2}$	Observed — Computed +	Observed — Computed −
(1)	(2)	(3)	(4)	(5)
33–37	73	5	68	...
38–42	119	21	98	...
43–47	207	89	118	...
48–52	421	266	155	...
53–57	599	587	12	...
58–62	957	954	3	...
63–67	1,147	1,142	5	...
68–72	982	1,007	...	25
73–77	660	655	5	...
78–82	252	313	...	61
83–87	72	109	...	37
88–92	15	29	...	14
93–98	1	6	...	5

These particular age groups are selected as there appears to be a slight excess in the number of entrants at decennial

and quinquennial ages, and by placing these in the middle of the groups we get rid of the disturbance, which would otherwise affect the numbers.

An examination of the numbers in column (2), between ages 53 and 78, shows that they form a nearly symmetrical curve, as is seen by a comparison with a "normal" curve of frequency given in column (3).* The numbers above age 78, however, are in defect, and those below 53 are considerably in excess of the figures suggested by the normal curve. As regards the falling-off of the numbers at the older ages, it may be conjectured that it is in part due to the fact that many published tables of the cost of annuities cease at age 75 or 80. The observed excess in the number of entrants at ages below 50 evidently represents the entrance at these ages of a class of lives differing from those forming the bulk of the data. It may perhaps be conjectured that a number of these cases are counter lives in contingent reversions, or similar securities, upon whose lives annuities have been purchased to secure the payment of annual premiums. Be that as it may, we find that while the deficiency of entrants at the older ages does not appear to affect the mortality rates, the entrants at the younger ages on the contrary show abnormally heavy mortality, the ungraduated values of the expectation of life for entrants under age 55 being relatively low. Hence we may calculate that the male annuitant experience is heterogeneous, and in using the results as a basis of calculation for the future, the abnormal part of the experience representing the entrants at the younger ages was properly rejected.

In addition to tables of the kind we have been considering, a statistical table may be a numerical statement of the manner in which variation in one particular from the average of the group is accompanied by variation in some other particular. We may, for instance, have a table representing a number of individuals, arranged according to height, the numbers at each height being further arranged according to weight. We should then have a table of double entry, each row or column of which would represent a statistical table of the form already considered. By means of this table we should be able to "correlate", as it is termed, variations in respect to

* The constants of this curve were only roughly determined, but the agreement with the observed numbers between ages 53 and 78 is sufficiently close to illustrate the point under discussion.

weight with variations in respect to height. Such a table would represent a mass of figures, the bearing of which could not easily be grasped without some further analysis. If, however, we add to the table a column showing the average weight for persons of a given height, we then have a ready means of seeing how this average weight is affected by a change in height. Having inserted the average, we have not exhausted the information which the original figures give us. We need also to know to what extent on the average the weight varies when the height remains constant; that is, we need to insert against each average weight what we have termed the "standard deviation."

A familiar example of such a table is one showing the ages of husbands and wives at marriage. Such a table would take the following form—

TABLE IV.

Showing Ages of Husbands and Wives at date of Marriage.

Husbands' Ages	Wives' Ages: under 20	20–30	30–40	40–50	50–60	60–70	Mean Ages of Wives
under 20	13	5	...	...	...	...	17·8
20–30	215	500	16	1	...	...	22·3
30–40	14	107	39	4	...	...	27·0
40–50	1	14	23	12	2	...	35·0
50–60	...	2	6	9	4	...	42·1
60–70	...	...	1	3	4	2	52·0
70–80	...	...	...	1	1	1	55·0
Mean Ages of Husbands	25·1	27·2	37·6	49·0	58·6	68·3	—

If there were no correlation between the ages of the husbands and wives at marriage, the figures showing the average ages for the various columns would (except for accidental fluctuations) be identical, and the same would hold for the average ages of the successive rows.

If a line were drawn through the table cutting those points in the *rows* corresponding to the average ages, and another line similarly cutting those points in the *columns* representing average ages, it would be found that these points could roughly be represented by straight lines, which

in the present example would be nearly coincident, since the spread of the figures, as measured by their standard deviation, is very similar in both rows and columns.

It is not always the case, however, that the nature of the correlation can be represented by a straight line. In the following example we have a somewhat different class of table showing the proportions for different age groups of wives and widows in an Indian pension fund.

TABLE IVA.

Showing proportion of Wives and Widows in a Pension Fund.

Ages	Number of Wives	Number of Widows	Total	Widows, per-cent of Total
under 20	19	...	19	0·0
20–30	1,430	50	1,480	3·4
30–40	3,366	355	3,721	9·5
40–50	3,329	1,018	4,347	23·4
50–60	1,653	1,312	2,965	44·2
60–70	476	933	1,409	66·2
70–80	63	330	393	84·0
80–90	6	46	52	88·5

Here it will be seen, from the run of the figures in the last column, that they cannot be well represented by a straight line, being somewhat in the form of the curve of $\int e^{-x^2/c^2}dx$, or of the curve $\frac{a^x}{m+a^x}$, with values of 0 and 1 respectively at the limits.

Such a table of correlation has an analogy with the table of the "Exposed to Risk" and "Died", which ordinarily forms the basis of our Mortality Tables. This table is virtually in the following form—column (4) representing the number of annual survivors being usually omitted as being implicitly contained in columns (2) and (3)—

Table of Exposed to Risk and Died.

Age	Exposed to Risk	Died	Survived
(1)	(2)	(3)	(4)

We have here the ages of the persons observed; the numbers under observation, or "Exposed to Risk", which, for the sake of simplicity, we will suppose to remain under observation for the entire year of age; the number of those who die during the year, and of those surviving. If we represent the rate of mortality by q_x then in all cases in column (3) $q_x=1$, and in all cases in column (4) $q_x=0$, and we have a table which is analogous to the table of the weights of individuals of respective heights, only that instead of having various values of q_x, we have in the nature of things only two possible values 0 and 1, the average value for each group representing the observed "rate of mortality." This table differs from that correlating weights and heights, or ages of husbands or wives at marriage, agreeing with that correlating age and civil condition, in the fact that a certain quality or characteristic, in this case death during a given year of age, is not present in varying proportions, but is either present or entirely absent. We are thus introduced to the conception of probability, the proportion of any group surviving or dying representing the "probability" of survival or death for any individual of the group taken at random. The idea of probability is also present in the supposed table of weights, although not so obviously. That table would inform us, for example, of the probability of a person of given height exceeding or falling short of a certain fixed standard weight, and we should then have a table identical in form with the table of Exposed to Risk and Died.

This conception of probability is important to the Actuary, because his object in collecting statistics is the distinctly practical one of measuring the probability of the happening of certain contingencies. It is necessary to realise clearly what is meant by the statement that the probability of a particular event has this or that value. Laplace pointed out that when we speak of the probability of the happening of a given event, we do so only on account of our ignorance of the antecedents of the event, or our inability to completely analyze them. If we entirely knew the antecedents, and if our powers of analysis were equal to the task, we could predict the event. In many cases we are able to do this approximately, but where the effective causes at work are numerous and obscure, and the result in individual

(apparently similar) cases is very variable, as in all questions affecting life contingencies, we are unable to forecast the event in a given case, and must fall back upon the average result deduced from the examination of a large number of similar cases. In other words, we treat the particular case in question as one of an indefinitely large class of similar cases, a sample of which we have already had under examination. From the results of such examination we infer the composition of the class as a whole, and hence the "probability" or average event in an individual case. If, in the sample observed, a given character is present in a certain proportion of cases, as for instance, where out of a number of persons of given age under observation, a certain proportion have died within the year of age, then we estimate the probability of the event happening in a particular instance, by the ratio which the number of cases in which the event has occurred bears to the entire number of cases observed.* To determine the probability of a given event is therefore to assign the case to the natural group or series to which it properly belongs and to pass under examination a sample of the group sufficiently large to enable us to determine approximately the average character of the whole as regards the particular quality in question. We are here speaking of simple events; the probability of a complex event, such as the survival of one life by another, is, of course, not determined directly by past observations. The latter yield the simple probabilities of surviving each year of age, by suitably combining which we arrive at the value of the probability desired.

The degree of certainty with which we can deduce the properties of an entire class from the part known to us, depends first on our assurance that the class is homogeneous, or at least that the portion observed is representative, such as would result from a selection of cases made at random, and secondly, on the number of cases that have been under

* The formula deduced by Laplace by which the true probability of an event which has been observed to happen m times out of $m+n$ trials is taken as $\frac{m+1}{m+n+2}$ is obviously not applicable to such a function as the rate of mortality, nor to any analogous function. It is sufficient to consider that in tabulating the values of the probability of dying in each year of age, we are using an arbitrary unit of time which might just as well be a month or day, in which cases we should, by use of the above formulæ, produce quite different mortality tables from the same data.

observation. If we examine the figures in tables similar to Tables I and II, we see that, in proportion as the number of cases under observation is small, the figures representing the results of the experience are irregular, while, on the other hand, where the number of facts observed is very large, the irregularities become relatively less. We arrive at the same conclusion from theory. If an indefinitely large group N contains Np objects of class A and $N(1-p)$ objects not of class A, and if from the group n objects are selected at random, then on the average np of these will be of class A. If we represent the observed number in any given case as $np+z$, the average algebraical value of z will be zero, while its average numerical value, irrespective of sign, will be very nearly $\frac{4}{5}\sqrt{np(1-p)}$.* This latter quantity clearly increases as np increases, but at the same time its ratio to np diminishes. Thus in a table of exposed to risk and died the actual irregularities in the number of deaths increase with the magnitude of the experience, but the irregularities in the rate of mortality diminish. Hence from theory as from experience we derive the conviction that if instead of the limited number of facts which we have been able to examine, we could have examined an indefinitely large number of similar facts, the results would have been relatively free from irregularity, and capable of being expressed by a continuous curve; without, of course, being sure that any such curve could be expressed algebraically.

The idea underlying the graduation of the figures of a statistical table, whatever be the process employed, is that a continuous curve may be found representing the general trend of the observations freed from irregularities due to paucity of material. This curve, we have reason to believe, will correspond more closely than the ungraduated curve to the results obtainable from a much larger body of facts. This is the rationale of the process of graduation and its justification. Such a process cannot deal with systematic errors affecting the table as a whole and cannot compensate for inadequate data. It adds weight to the results, however, at each individual point of the table, and assists in bringing into relief the true character of the curve by freeing it, in a large measure, from accidental irregularities.

* The average value of z^2 will be npq, the average value of z^3 will be $npq(p-q)$, and the average value of z^4 will be $npq[(3n-6)pq+1]$. See Note A, p. 110.

There may be other objects aimed at in a graduation besides that of removing the irregularities from the rough figures, with the view of bringing out more clearly the law underlying them. The Actuary constructs tables not merely to show what has happened in the past, but to enable him to forecast the future, and as he requires these tables as a basis for financial operations, considerations are introduced which do not arise in the treatment of purely statistical tables. Whatever class of events the Actuary may have to deal with, will be subject to change with the lapse of time. That portion of the class he has been able to observe lies necessarily in the past; the conclusions he has derived from their study he proposes to extend to the future. He must therefore consider how far the observed characters of the class are changing or permanent, and must endeavour to distinguish between changes representing permanent tendencies and those due merely to temporary fluctuations. In the selection of data suitable for his purpose the Actuary will aim on the one hand at a sufficiently broad basis both in space and time to eliminate the effects of local and temporary fluctuations, and on the other hand he will aim at obtaining as far as possible a homogeneous group of data. These two aims are more or less in conflict, and he will lean to the one side or the other, according to the object he has in view. Where, for example, that object is to produce a table that may be adopted as a general standard by various institutions, often differing considerably as to their individual experience, he must aim at a correspondingly broad foundation. In these circumstances it will not generally be possible to obtain a really homogeneous experience. If it is a question of the mortality of assured lives, for instance, this will be found to be affected by endless individual variations, age, sex, duration of assurance, occupation, civil condition, class of assurance, character of the insuring office, &c., &c., and from such material approximately homogeneous data could only be obtained by cutting up the experience into comparatively small groups and thus sacrificing all generality. This can be avoided in practice by first excluding all extreme variations. The sexes will be separately treated, lives so impaired as to prospects of longevity by personal health, family history, occupation, or residence in unhealthy districts as to be "rated up" will be excluded, as also classes of assurance that may

be supposed subject to rates of mortality differing from the average. When the data has thus been trimmed of the extreme variations, a body of experience will generally remain not greatly shrunken from its original dimensions and in which the discontinuous variations are sufficiently numerous and individually unimportant to render the data for practical purposes homogeneous. The rates of mortality, or of withdrawal, can then be treated as functions of the two remaining variables of importance, the age and the time elapsed from date of entry; or as functions of the age only from the point at which the factor of duration may be found to be unimportant.

On the other hand, the Actuary's object may be precision rather than generality; he may have to deal with a group, subject to special conditions and presenting special characteristics, as is usual in the case of pension funds and friendly societies. Here, if the data are at all adequate, better results will be obtained therefrom than by having recourse to any general experience. Where it is insufficient by itself as a basis for statistical tables it may serve as an indication as to what standard table is the most suitable to employ and as to how far and in what direction it may be desirable to introduce any modifications therein. In an experience of this character the data may sometimes be very heterogeneous, but there is usually the safeguard that its composition is approximately constant.

A question of some importance may here be considered, namely, the relative claims of lives, policies, or amounts assured to form the basis of the mortality table. In the 17 Offices' data, the number of policies, in the H^M and O^M data, the number of lives passing under observation constitute the basis of the experience, while in the American Offices' Experience (1880) the sum assured was the unit. In the instances of the H^M and O^M Tables, wherever a life would have been doubly observed the duplicate assurance was eliminated. In justification of the use of the sums assured as the basis of the experience, in lieu of the number of lives, it may be said that in this way we represent the financial effect of the mortality, as it makes no difference to the insuring company whether one claim arises for £10,000 or one hundred claims for £100 each. There are, however, serious objections to employing the sums assured as a basis for a

mortality table, based upon a general experience. Either the mortality among the lives carrying large sums assured is similar to the average or it is not. If it is similar, the general character of the table will not be affected by the additional weight given to these lives in the experience, but the irregularities in the deduced rates of mortality will be considerably increased. The result, indeed, will be virtually the same as if we had used a part only of the available data, selected at random, instead of the whole. If, on the other hand, the mortality among the lives insured for large sums is materially different from the average, then the experience is not homogeneous. As a matter of fact, these lives of themselves do not form a homogeneous group. In certain societies they appear to give better rates of mortality than the average; in others, where they are mainly represented by non-profit policies effected for commercial reasons, they are no doubt subject to higher rates of mortality than the average. As in a general experience, combining the individual experience of many offices, these lives will represent an exceptional or abnormal element, which may or may not persist in the future, and will certainly not persist equally in all societies, it is not desirable in deducing a general mortality table to specially "weight up" this part of the data.

The same considerations apply, but with somewhat less force, to the plan of making policies rather than lives the basis of an experience. Without dogmatizing upon the point, it appears to me that the proper course is, where two or more policies are effected at the same time or at the same age at entry, to treat them as a single risk, but where the subsequent policies are effected at later ages, involving fresh medical selection, to treat them as separate risks. This means the elimination of duplicates in each of the "select" tables for individual ages at entry, but no further elimination in the resulting aggregate tables, a course which has the advantage of making the aggregate table the true aggregate of the tables for separate ages at entry. Judging by the results of the O^{M} experience, this course is necessary if we are to produce an aggregate table, representing "ultimate" rates of mortality after the lapse of a stated period from entry, which will join on smoothly to the "select" rates.

A detail of less importance, but of considerable interest, is the question of the proper treatment of withdrawals in a mortality experience. These are usually treated as withdrawing upon the termination of the days of grace in case of lapse by non-payment of premium, and for the purpose of obtaining the true measure of the mortality experienced this course is the correct one. It should be borne in mind, however, that to arrive at the financial effect of the mortality the numbers of the exposed to risk should correspond to the number of annual premiums paid, and from this point of view the life withdrawing should not be treated at risk during the days of grace. The differences in the resulting mortality rates according to the two methods is, of course, very slight.

SECOND LECTURE.

HAVING dealt in the last lecture with the *rationale* of graduation in general, I now propose to refer more particularly to the principles underlying certain special methods of graduation. We may divide the various methods which are in use into three classes:

1. Graphic methods.
2. Methods based upon Interpolation or Finite Difference formulæ, such as Mr. Woolhouse's.
3. Methods which depend upon the use of Frequency Curves, in which we may include all methods based upon the assumption that the series to be graduated can be represented as some function of the variable.

Certain general considerations apply to all these methods. We may have to deal either with *a single series of numbers*, such as the number, at successive ages, of lives effecting assurances, of persons enumerated at a census, or attacks from a given disease, &c.; or, as more often happens in actuarial statistics, the fact of importance may be *the ratio between the corresponding members of two series of numbers* as in a table of "Exposed to Risk" and "Died", forming the basis of the Mortality Table, where the fact sought is the rate of mortality at each age given by the ratio of the Died to the Exposed to Risk, the actual numbers of these being of importance mainly as affording a measure of the trustworthiness of the deduced ratio.

Where only a single series of numbers is involved, the problem is comparatively simple, and an accurate solution is not generally of great importance to the actuary. In the more usual case where the ratio of the corresponding members of two series of numbers is in question, the problem is more complicated. We have a choice of procedure: we may either graduate independently the two series of numbers

(in the case supposed the numbers of the "Exposed to Risk" at each age and the numbers of the "Died"), or, disregarding the irregularities in the two series, we may proceed to deal at once with the ratios only. If each series can be satisfactorily graduated, the resulting curves being smooth and fitting the ungraduated series sufficiently closely—that is to say, within the limits of the errors of observation—we may then assume that the ratios of the corresponding terms (in the case supposed the rates of mortality) will also be within the limits of error. It may also be said that by working with the rough facts themselves, rather than the ratios between the two, we keep in view the weight of the observations at each point of the curve, and are able to see at once how far our graduated numbers vary from the original, and how far that variation is justified by the number of facts at each particular point. There are, however, some important objections to this course. In the first place, the ratio between the corresponding terms in the two series of numbers represents generally a relatively stable quantity, whereas the actual numbers in either series, depending as they do upon the extent of the experience under review at particular ages, are liable to fluctuations of a more or less arbitrary character. Further, supposing the graphic method of graduation or the method of finite differences is employed—in either case the argument is applicable, although specially so in the former—it will be found that each curve will contain certain outstanding irregularities, as it is not possible entirely to remove all irregularities by those methods. Hence in the adjusted ratios two sets of irregularities will be superimposed and a less satisfactory series of values obtained than if the ratios themselves had been dealt with.

A stronger objection, when dealing with a mortality experience, to graduating separately the numbers in the two series of "Exposed to Risk" and "Died" rather than their ratio, is that we thereby discard our previous knowledge of the nature of the curve expressing that ratio—our general knowledge, that is, of the nature of the curve q_x or μ_x—knowledge which is of considerable assistance in graduating the commencement and end of the table where the data are few.

Where a graduation of both series of numbers is made, it is preferable, indeed necessary if the best results are to be obtained, after first graduating the series corresponding to

the "Exposed to Risk", to re-compute the numbers of deaths, lapses or marriages, as the case may be, on the basis of the graduated numbers of the Exposures, and to operate upon these adjusted numbers. We are in this way less likely to obscure the law of the series representing the required ratios.

Notwithstanding any theoretical objections, there may be occasions on which it is more convenient, or even necessary, to deal with the two series separately; where, for example, as in the Registrar-General's returns of the population and deaths for certain occupations, we have not the facts for individual ages, but only in certain large groups. The ratio of deaths to exposures for each age group are obviously not satisfactory approximations to the rate of mortality for the central age of the group. In these circumstances it appears to be best to adopt a plan similar in principle, though not in detail, to that employed by Milne in graduating the Carlisle Table, and to draw curves respectively through the parallelograms representing the exposures and the deaths, and from these deduce the numbers for individual ages. The graphic method, however, is not very suitable for this purpose, and the use of interpolation formulæ does not always give good results. It is generally better to make use of suitable frequency curves. It will be seen later that, where the number of groups is rather small, the use of the normal frequency curve, with certain modifications, enables us to re-distribute the numbers representing the groups of "Exposed" and "Died", and so obtain graduated numbers for each age, and hence from the ratios of these a graduated rate of mortality. (*See* the Sixth Lecture, p. 91.)

We shall now assume that we are dealing, not with the two independent series, but with the ratio between the two; as, for example, with q_x, or some analogous function.

We may consider we have three independent estimates of the value of q_x:—

1st—That derived from the observed ratio of the died to the exposed at age x.

2nd—That derived from the data at neighbouring ages.

3rd—That derived from previous experience of more or less similar data.

The first and second should be suitably combined in the process of graduation. The last is, in the nature of things, a very vague estimate, and bears a relation to that derived directly from the observations, if these are numerous, similar to that of a rough measurement by inferior instrumental means to one made by an instrument of precision. In such case no weight attaches to it.

There are circumstances, however, in which the *a priori* estimate of the values of q_x become important, viz., when the observations at our disposal are extremely few. As the extent of our observations diminish, the numbers of exposures and deaths becoming smaller, the weight to be attached to the deduced values of the rate of mortality become less, and a point is eventually arrived at when we obtain more trustworthy results by considering to what particular class of examined data the experience most nearly conforms in character, and falling back upon the results of such related experience.

If we have to deal with a large experience, a somewhat similar difficulty arises at the commencement and end of the table. Generally speaking, we then derive more trustworthy values for the rates at these ages from a consideration of the general trend of the curve and our previous approximate knowledge of its character, than by falling back upon any related experience.

Coming to the principles underlying each of these three methods of graduation, we consider first the graphic method, whether in the form employed by Milne or in the preferable form employed by Dr. Sprague. This method makes no further assumption than that the series with which we are dealing would, if the observations were sufficiently extensive, form a continuous and regular curve, and that the irregularities actually occurring in the ungraduated values are due to the smallness of the data.

To Dr. Sprague (*J.I.A.*, vol. xxvi, p. 77) we owe the most systematic and satisfactory exposition of the graphic method. An essential feature in his procedure is the preliminary division of the data (which we may suppose arranged by years of age) into groups, so selected as to afford a steady progression in the average rates of mortality for

successive groups, due regard being had to the range of these groups. For examples of the method, the student must be referred to Dr. Sprague's original papers. This process of dividing the data into selected groups appears at first sight to be arbitrary, but it may be justified on the grounds :—(1) That in a series of observations such as we are discussing, where at each age the results are affected by irregularities or errors of observation, a successful graduation will reduce the sum of these errors and also the sum of the "accumulated" errors to zero, or nearly so. Hence if we compute at each age the accumulated errors (reckoning from either end of the series) these must, in order that their sum may be approximately zero, change sign, thus passing through zero, fairly frequently. The data will, therefore, be made up of consecutive groups, larger or smaller, in each of which there is an approximate balance of errors, and it may be assumed that, with a sufficient amount of experience and the exercise of some trouble, these groups can be found by inspection and trial. (2) In further justification of this procedure, it is to be noted that the rates of mortality deduced from the average rates in the selected groups are used as a first approximation only, the final rates being arrived at by repeated comparison of the graduated deaths with the actual numbers until a sufficiently smooth curve and a sufficiently close agreement has been obtained. At the same time I am not convinced that the use of these specially selected groups has any real advantage over the use of groups of constant range, as quinquennial or decennial, provided the operator recognizes that he cannot look for an absolute balance of errors in these latter, but must regard them as equally subject to errors of observation with the numbers at individual ages.

Assuming it to be practicable to draw a sufficiently smooth curve, free from sudden changes of curvature, and yet representing the observations sufficiently closely with a due regard to their weight in different parts of the table, there would appear to be nothing to object to in the principle of the graphic method of graduation. In practice, however, there are certain difficulties. The first, particularly in the case of a mortality table, is the question of scale. Anyone who has attempted to make graphic graduations will, I think, have met with this practical difficulty. Whether we graduate separately the "Exposed to Risk" and "Died", or whether

we graduate a function such as q_x, the difficulty equally arises. The values of q_x may range in practice from about ·005 to, say, about ·5, and at the older ages increase so rapidly that the eye does not readily grasp the nature of the curve. In order that it may do so, and that the curve may be drawn and read off with sufficient accuracy, a certain proportion must be maintained between the horizontal and the perpendicular scale, so that the curve shall not cut the ordinates at too acute an angle. It is also necessary to represent the values of q_x in two or three sections, as the scale suitable to the older ages will not permit of the values at the younger ages being represented with sufficient accuracy.

Instead of operating on the rates of mortality, we may with advantage employ the logarithms of the rates, or the logarithms of the central death rates.* We thus obtain a curve which is much more easily dealt with. From the fact that the rates of mortality change slowly at the younger ages, and at the older ages generally approximate to a geometrical progression, the logarithms of the rates are nearly in the form of an arithmetical progression, and are represented by a line having very little curvature. At the oldest ages, indeed, it may very conveniently be taken as a straight line.

Perhaps the main difficulty in graphic graduation is that it is by no means easy, even with mechanical aids, to draw a sufficiently smooth curve. The curve as drawn may appear to be smooth, but on reading it off and examining the series of values obtained, we find irregularities which, in order to produce a satisfactory graduation, must be removed by a further adjustment. If we are dealing with a relatively small experience—in which cases these practical difficulties are correspondingly increased—they may be overcome to a large extent by using as a base line a well-graduated standard table representing an experience of similar character. By computing the "expected" deaths according to the standard table, and dealing with the ratio of the actual to the "expected" deaths in successive age groups, we avoid the difficulties due to inequality of scale and to the rapid increase in the value of the ordinates at the extreme ages. The curve

* *See*, however, Note B, p. 114, as to precautions in dealing with logs of rates of mortality and similar functions.

of ratios, apart from accidental fluctuations, will often be found to approximate to a straight line, the departures from which can be, of course, represented on a relatively large scale. In particular, the difficulty arising from the paucity of observations at either end of the table will be avoided by making each extremity of the curve of ratios terminate in a straight line, the locus of which will depend upon the general trend of the curve in the neighbourhood. The resulting values at the extremes of the table obtained in this way will be more trustworthy than those obtained without the aid of the standard base line.*

In finite difference or interpolation methods of graduation (of which we may take Woolhouse's as the best known type) the underlying assumption is virtually the same as in the graphic method, viz., that the curve is of such a nature that the ordinary methods of interpolation can be applied. Put more precisely, Woolhouse's method assumes that for a range of 15 consecutive ages the values of l_x can be represented with sufficient accuracy by a curve of the third order, *i.e.*, $l_{x\pm t}=l_x\pm at+bt^2\pm ct^3$ when t is not numerically >7. As this assumes the fourth and higher differences of l_x to be zero, we may write

$$l'_x=\frac{1}{125}\{-3(l_{x-7}+l_{x+7})-2(l_{x-6}+l_{x+6})+3(l_{x-4}+l_{x+4})$$
$$+7(l_{x-3}+l_{x+3})+21(l_{x-2}+l_{x+2})+24(l_{x-1}+l_{x+1})+25l_x\}$$

where l'_x may be taken as the graduated value of that function, the quantities on the right-hand side of the equation being the ungraduated values.

This formula, which is that used by Woolhouse in the graduation of the H^{M} Table, is of course only one of numerous possible formulæ deducible from the above expression for l_{x+t}. Others may be found resulting in a smoother graduated series, but all the formulæ since proposed as improvements on his are based upon the same general principle. An indefinite number of such formulæ can be found, even when the range is fixed.† In particular may be

* *See* Lidstone, *J.I.A.*, xxx, p. 212. These remarks are equally applicable to graduation by a finite difference formula (*see J.I.A.*, vol. xli, p. 89).

† *See* Todhunter, *J.I.A.*, xxxii, 378; G. F. Hardy, *J.I.A.*, xxxii, 371.

mentioned Mr. J. A. Higham's, Dr. Karup's, and that used by Mr. J. Spencer in the graduation of the "Manchester Unity" mortality experience. *See* the following table showing the value of u'_x in terms of the ungraduated u's:—

TABLE V.

Showing the values of ϕ_t, where $u'_0=\Sigma_{-t}^{+t} u_t \times \phi_t$, by various well-known Graduation Formulæ.

Distance from Central Term t	Spencer 21-term Formula	Karup	Higham	Woolhouse
0	·172	·200	·200	·200
± 1	·163	·182	·192	·192
± 2	·135	·139	·144	·168
± 3	·095	·085	·080	·056
± 4	·052	·034	·024	·024
± 5	·017	·000	·000	·000
± 6	−·005	−·013	−·016	−·016
± 7	−·015	−·014	−·016	−·024
± 8	−·015	−·010	−·008	·000
± 9	−·009	−·003	·000	⋮
±10	−·003	·000	⋮	⋮
±11	·000	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮

It is clear that no such formula will entirely remove the irregularities in the series, and in Woolhouse's graduation of the H^M Table the outstanding irregularities were removed by an empirical process similar to that employed for the graduation of the 17 Offices' Table, and described in his paper (*J.I.A.*, vol. xii, p. 140-1). The object aimed at in a formula such as these, should be so to select the coefficients of the terms on the right hand that, while giving an expression for the value of the central function correct as far as the order of differences employed, the formula will produce the maximum smoothness in the flow of the graduated values. This may be done by simple experiment, or we may adopt some empirical measure or standard of smoothness and thereby compute the most advantageous coefficients. We may, for example, adopt as our standard of smoothness the extent to which the second differences of our graduated function are affected by the errors of observation in the original table.

Applying this standard to Woolhouse's formula, we have

for the graduated second central difference of l_x (using central differences for the sake of symmetry)—

$$\begin{aligned}125\Delta^2 l'_{x-1} = &-3l_{x-8}+4l_{x-7}+l_{x-6}+l_{x-5}+l_{x-4}+10l_{x-3} \\ &-11l_{x-2}-2l_{x-1}-2l_x-2l_{x+1}-11l_{x+2} \\ &+10l_{x+3}+l_{x+4}+l_{x+5}+l_{x+6}+4l_{x+7}-3l_{x+8}.\end{aligned}$$

If we assume that on the average each of the ungraduated values of l_x on the right-hand side of this equation is subject to a mean error of $\pm e$, and if we assume that these errors may be combined according to the normal law, then the mean error of the entire expression for $\Delta^2 l'_{x-1}$ will be found by multiplying e by the square root of the sum of the squares of the coefficients, giving

$$\frac{(\sqrt{3^2+4^2+1^2+1^2+1^2+\ \&c.})}{125}\, e = \frac{\sqrt{510}}{125}\, e = \cdot 18e.$$

In the same way it may be shown that in Karup's formula the mean error in $\Delta^2 u_{x-1}$ is about $\cdot 068e$, where e is the mean error of a single value of u_x. It must not be supposed from these results that the mean errors in the graduated values of l_x or u_x are proportionately reduced. The mean errors in the graduated functions when Woolhouse's formula is employed are reduced to about ·42 of the mean errors in the ungraduated functions, or are about equivalent to the mean errors of the ungraduated values corresponding to an experience $5\frac{1}{2}$ times larger. The graduated table based on the smaller data would, however, be *smoother* than the ungraduated table based upon the larger data. (See *J.I.A.*, xxxii, pp. 376-7.)

Taking a generalized formula, such as

$$u'_x = au_x + b(u_{x-1}+u_{x+1}) + c(u_{x-2}+u_{x+2}) + \&c. \ldots k(u_{x-t}+u_{x+t})$$

where u'_x represents the graduated value of u_x, and assuming that each of the ungraduated values u_x, &c., are affected by the same mean error $\pm e$, it is of course possible to determine the values of a, b, c, &c., so that the mean error in, say, $\Delta^2 u'_{x-1}$ shall be a minimum. Noting that $a=1-2b-2c-\&c.$, and that $b+4c+9d+\&c.=0$, in order that the formula may be correct to 3rd differences, an expression

may be found for $\Delta^2 u'_{x-1}$ in terms of u_{x-t-1}, u_{x-t}, &c., with coefficients involving $c, d, \ldots k$. If the coefficients of each term are now equated to zero, there will be $(2t+3)$ equations of condition with $(t-1)$ unknowns, which may be solved by the usual method of least squares.

This is somewhat theoretical, however, as the values we should obtain for the coefficients would be generally fractional, and the resulting graduation formula would not lend itself to any continuous method of computation, as is the case with Woolhouse's and other similar formulæ. An alternative would be to fix upon a convenient set of summations, and then to determine the function summed (called by Mr. Lidstone the "operand") so that (1) first and second differences may vanish—see *J.I.A.*, xxxii, 371, &c.; (2) The range of the formula may be what we require; and (3) that subject to (1) and (2) the coefficients shall be such as to make the mean error in Δ^2 or Δ^3 a minimum. This might give a fairly convenient working formula, as when once the operand was formed the ordinary convenient method of summation would apply.

If we consider the effect of such a formula of graduation upon the outstanding or unbalanced errors of observation in a small group of ages, we shall see that they are not very materially diminished. If, for example, we express the sum of five consecutive graduated values in terms of the ungraduated values, we shall have, in the case of Woolhouse's formula,

$$l'_{x-2}+l'_{x-1}+l'_x+l'_{x+1}+l'_{x+2}=\frac{1}{125}(80l_{x-2}+101l_{x-1}$$

$$+115l_x+101l_{x+1}+80l_{x+2})$$

$$+\text{terms involving other values of } l.$$

Here it is obvious that any systematic or unbalanced error in the original group will not be greatly reduced (probably to about three-fourths of its amount) in the graduated table. While, therefore, finite difference formulæ of graduation yield, generally, a smooth curve as regards the progression of the graduated values from age to age, they have a tendency to reproduce any waviness in the original, due to the unbalanced errors affecting small groups of four or five consecutive ages.

A question arises in connection with this method as to what particular function should be selected for graduation. In the case of Woolhouse's original formula the function operated upon was l_x. Practically speaking, except for the latter portion of the table, this approximates in result to a graduation of the rates of mortality. This may be seen from the following relations. Any adjustment of the l_x column by a finite difference formula has, of course, the same effect as a similar graduation of the d_x column. Since $d_x = l_x q_x$, and since for the range of ages included in the formula (fifteen in Woolhouse's formula, of which, however, only the five central ages are heavily weighted) the values of l_x are not in general widely different, the graduation of the l_x or d_x column should give results not materially different from those obtained by graduating q_x. At the older ages, however, there may be significant differences in the results, and I must express my preference for the rate of mortality as the more suitable function to graduate if the observations are duly weighted or if proper precautions are taken to avoid anomalous results at either end of the table where data are scanty.

An objection to the principle of the finite difference methods of graduation is that the weight of the observations is not allowed for at various ages. This objection is not very serious, however, as at the commencement and end of the table, where it would be chiefly felt, the method is usually not strictly applied. It may be noted that if the l_x function be graduated, then its rapid decrease in value at the oldest ages in the table gives automatically a diminishing weight to the observations with increasing age, but at the same time yields somewhat irregular graduated values. The objection may, of course, be got rid of by first applying a smooth series of weights to the function to be graduated, prior to graduation, and eliminating these factors afterwards.

A difficulty arises in the use of finite difference formulæ from the smallness of the data at the extremes of the table and from the fact that the first 7 or 8 values of the graduated function cannot be obtained from the formula. In the case of a mortality table there is not so much difficulty in dealing with extreme old age, because there, as Woolhouse points out, if we are dealing with the function l_x it may be taken $=0$ beyond the limiting age of the table, or if

we are graduating the rate of mortality, q_x may be put down as equal to unity. As regards the earlier ages, Woolhouse's method is to obtain from the formula the graduated values of l_x so far as this can be done, that is, to within 7 years of the initial age, and to compute the values for the first seven ages of the table from the values of l_0, l_7, l_8 and l_9 (l_0 representing the value of l_x for the initial age) on the assumption of a constant third difference. This method may in certain cases lead to anomalous results, even negative rates of mortality. Mr. Ackland has given an alternative method of considerable ingenuity (*J.I.A.*, vol. xxiii, p. 357). The difficulty may be avoided by assuming values for the initial ages, as, for example, a constant average value of q_x or d_x, or other arbitrary values deducible from the general character of the experience. A more satisfactory method would be to determine q_x for the first 10 or 15 ages, by the method of moments or least squares, on the assumption that it could be represented by a first or second difference function. All these methods, however, are expedients more or less empirical, though they may in practice lead to sufficiently satisfactory results.

The Finite Difference methods of graduation all assume that the functions to be graduated may be represented for successive small tracts of ages by a parabolic curve of the form—

$$u_x = a + bx + cx^2 + \text{ \&c.}$$

We are not bound to assume this particular form of function. We can employ the principle of the Interpolation method, representing our function by some other form, as, for example, $m_x = a + bc^x$ corresponding to Makeham's formula.

The principle of the methods of graduation we have been discussing, of which Woolhouse's is a type, must not be confounded with that used by Davies in graduating the Equitable experience, nor with that used by Mr. Berridge in graduating the Peerage mortality. These latter are more nearly allied to graduation by frequency curves than to Woolhouse's method. In Davies' Equitable graduation, curves of the third order are actually fitted to successive sections of the l_x column, the values of l_x from 10 to 40 being virtually found by a third difference interpolation from the

values l_{10}, l_{20}, l_{30}, l_{40}, those from l_{40} to l_{70} similarly from the values of l_{40}, l_{50}, l_{60}, l_{70}, and so on. Mr. Berridge's graduation of the Peerage mortality followed a similar principle, except that he represented the entire series of values of $\log l_x$ from 15 to 75 by means of a single curve of the sixth order, based upon the values of that function for decennial intervals of age.

As to the relative merits of graphic and finite difference methods of graduation, the former has an undoubted advantage when the number of facts at our disposal are few. In these cases formulæ of the type of Woolhouse's cannot be expected to produce very satisfactory results, as in the comparatively small section of the curve embraced by the formula the true character of the curve will frequently be obscured by the errors of observation. These formulæ are at their best when applied to a table based upon fairly extensive data, and presenting a curve without any rapid change of character. The advantages possessed by the graphic method in dealing with a small experience, owing to its flexibility and its power of bringing under contribution large sections of the curve at once, are, however, still more noticeable when frequency curves can be suitably employed.

We have already spoken of the success or sufficiency of a graduation, but we have not said anything as to what is the proper test of a successful graduation. Before dealing with the general principle of graduation by means of frequency curves, it will be useful to consider this question. There are obviously two conditions that should be fulfilled by a graduation. In the first place, a smooth and continuous progression in the graduated values. This is required, because we have good reason for believing that if the true values were ascertainable, they would exhibit this property. In the second place we require an adherence to the original data, sufficiently close to be fairly within what we may conveniently term the errors of observation.

The standard of smoothness is not easy to define. If a formula is adopted representing the ultimate values of l_x, q_x, or μ_x as a function of the age, this in itself secures a smooth series. In other cases the sufficiency or otherwise of the graduation in this respect must be left to individual judgment. The advantages of a really smooth curve are

mainly found where it is necessary to resort to interpolation or to the use of summation formulæ; and, further, in the practical consideration that with a really smooth curve nearly all tables calculated therefrom can be sufficiently checked by differencing.

As regards the second requirement, that of adherence to the general features of the ungraduated experience, it is easier to set up a criterion. We have already seen that if the true value of the probability of an event happening at a single trial is p, the event will, on the average, happen np times in n trials, and if there are series of n_1, n_2, n_3, &c., trials in which the probabilities of the respective events are p_1, p_2, p_3, &c., then *on the average* the total number of occurrences in such a series of trials will be $n_1p_1+n_2p_2+n_3p_3+$, &c. That is to say, if the observed occurrences are θ_1, θ_2, θ_3, &c., then the *average value* of each term $(\theta_1-n_1p_1)$, $(\theta_2-n_2p_2)$, &c., and consequently of the sum of such terms, will be zero.* It is also obvious that the *average value* of the sum of the series $(\theta_1-n_1p_1)+2(\theta_2-n_2p_2)+3(\theta_3-n_3p_3)+$, &c., and generally of the series whose rth term is

$$\frac{\underline{|r}}{\underline{|t}\,\underline{|t-r}}\,(\theta_r-n_rp_r)$$

will be zero. In the case of a mortality experience these quantities $(\theta_1-n_1p_1)$, &c., represent the deviations of the observed deaths at each age from the "Expected Deaths", as computed by the true rates of mortality, supposing these to be known. It follows, therefore, that we should expect the total of such deviations *on the average* to be zero, and in the same way the average value of the successive sums of the accumulated deviations should be zero. Generally, if we put

$$\Sigma n_0=n_0+n_1+n_2+n_3+, \text{ \&c.}$$

$$\Sigma\Sigma n_1=\Sigma^2 n_1=n_1+2n_2+3n_3+\quad, \text{ \&c.}$$

$$\Sigma\Sigma\Sigma n_2=\Sigma^3 n_2=n_2+3n_3+6n_4+, \text{ \&c.};$$

we shall have on the average

$$\Sigma^t(\theta_r-n_rp_r)=0.$$

* This is not the *most probable* value of these terms, although in general it will be very close thereto. The Actuary, however, requires to consider the *average* result, not the most probable.

We should not expect (assuming the true values of p_r to be known) that these sums of the deviations of the actual from the expected numbers would actually be equal to zero in any given case, but we should expect in a long series of cases that the positive values would approximately balance the negative. We do not expect to obtain exactly 1,000 heads in a series of 2,000 tossings of a coin, but we should expect to find that the average number of heads over a great number of such series of tossings would be very close to that figure. This reasoning leads us to the conclusion that, given a successful graduation, we should not only have obtained a smooth series, but that the sum of the deviations between the computed events (deaths or otherwise). and the observed numbers, would be nearly zero, and that the successive sums of the accumulated deviations would be small.

It is not necessary in practice that this test should be pushed too far. We may be satisfied if the sum of the deviations and the sum of the accumulated deviations are practically zero; if the total deviations in successive sections of the table (*e.g.*, in quinquennial or decennial groups) appear to be, on the whole, within the limits of the errors of observation; and if the total of the accumulated deviations changes sign fairly frequently. On the other hand we should expect that the total deviations irrespective of sign should not be *materially less* than their theoretical amount. Otherwise we should conclude that the series was under-adjusted and that accidental fluctuations in the curve had been incorporated as inherent characteristics.

These tests of a graduation are well known to Actuaries. and, indeed, have been very generally employed by them. So far as they go, they correspond to the method of moments which Prof. Karl Pearson has elaborated and employed with such success in the fitting of frequency curves to statistical data. It is clear, however, that they can only be employed systematically in conjunction with those or other curves capable of analytical expression. Using methods of graduation, based upon Finite Difference formulæ, such as Woolhouse's, we cannot secure that the successive sums of the deviations shall vanish, though in general we may expect them to be small. Using the graphic method, we can, by a gradual process of hand-polishing the curve, reduce the

accumulated deviations and their sum to as small a value as we please,* but the process is a tedious one.

A second test that has occasionally been applied when the graduation has been effected by means of a formula, is that of making the sums of the squares of the deviations a minimum, the deviations being either in respect of the graduated and observed deaths at each age or those of the graduated and ungraduated values of some function such as l_x or $\log. l_x$, This method, known as the method of "Least Squares", is used very generally in connection with measurements in astronomy and other physical sciences and has given rise to a quite extensive literature. It is based upon the assumption that if in a given series of observations the relative frequency of an error x at each observation is represented by the function ke^{-x^2/c^2}, then the probability of a conjunction of any

It may, perhaps, be worth pointing out that if we have obtained a smooth curve with a general conformity to the original facts, but not making the Σ (deviations) or Σ^2 (deviations) vanish, this may be done by the following plan. Assume, for the sake of illustration, that the function graduated is the central death rate m_x. Representing by m_x the graduated values of that function by E_x the "Exposed to Risk" in the middle of the year of age and by θ^x the observed deaths, let

$$\Sigma(m_x E_x - \theta_x) = A$$

$$\Sigma^2(m_x E_x - \theta_x) = B$$

then, if $m'_x = a + (1 + b)m_x$ be the modified rates required,

$$a \,.\, \Sigma(E_x) + b\Sigma(E_x m_x) = -A$$

$$a \,.\, \Sigma^2(E_x) + b\Sigma^2(E_x m_x) = -B$$

whence a and b are determined.

If the table on the whole follows Makeham's law the use of this form of correction enables us to neglect all orders of differences in the preliminary adjustment of m_x or μ_x. Formulæ may thus be employed (as for example, a simple double summation in groups of 10 values, or, still better, successive summations in 10's, 5's and 2's) giving a much smoother curve than when account has to be taken of second differences, the resulting systematic error of this first graduation being corrected as above.

In the alternative, if $m'_x = m_x + a + bx$,

$$a\Sigma(E_x) + b\Sigma x(E_x) = -A$$

$$a\Sigma^2(E_x) + b\Sigma^2 x(E_x) = -B.$$

This method may be employed in conjunction with Mr. Lidstone's plan of using a standard table as a base line for purposes of graduation.

set of errors x_1, x_2, x_3, &c., will be proportional to the value of the product

$$e^{-x_1^2/c^2} . e^{-x_2^2/c^2} . e^{-x_3^2/c^2}, \text{ \&c.,}$$

$$= e^{-\left(\frac{x_1^2+x_2^2+x_3^2+\ldots}{c^2}\right)}$$

which clearly has a maximum value when the index of e is numerically a minimum, *i.e.*, when the sum of the squares of the errors $(x_1^2+x_2^2+x_3^2+\text{\&c.})$ is the least possible. This expression assumes that the average error, and therefore the probability of a unit error, in each observation is the same, an assumption which may often be fairly made in respect to independent measurements of a physical quantity. If the observations are not of the same weight, so that the probability of the errors of x_1, x_2, x_3, &c., in the respective measures are

$$e^{-x_1^2/a^2}; \; e^{-x_2^2/b^2}; \; e^{-x_3^2/c^2}; \text{ \&c.,}$$

then the most probable solution will evidently be that which makes the sum of these exponents the least possible.*

The assumptions upon which this method is based are not strictly in accord with the conditions of a mortality experience or similar statistical observation. If the method is applied to the deviations between the observed and graduated deaths, the objection may be raised that the observations at different ages are not of equal weight, and that the probability of a unit error varies at each successive age, while in each case the probability of a given error can only be approximately expressed by the normal function ke^{-x^2/c^2}, positive and negative errors not being equally probable. It is, of course, possible suitably to weight the observations, so that a unit error is made equally probable. For example, if at any given age there are n "exposures", and if the true probability of death is q, then the "standard deviation" or $\sqrt{\text{average square deviation}} = \sqrt{nq(1-q)}$, and the probability of a difference of x between the expected and observed deaths is approximately $\kappa e^{-x^2/nq(1-q)}$; the error in the formula when x is positive nearly compensating the error when x is negative. Hence, if the "Exposed to Risk" and "Died" at each age are multiplied by the factor $[nq(1-q)]^{-\frac{1}{2}}$, *where q*

* *See* Note C, p. 117.

is to be taken at its true or graduated value,* then the observations may be considered to be properly weighted for the application of the method of least squares.

We shall see in the following lectures that there is an intimate relation between the criteria of least squares and moments. This will be better discussed after considering the question of frequency curves and the process of fitting them to a set of statistical observations.

The ungraduated values of q cannot be used, as this would result in undue weight being given at all ages where the observed mortality was in excess of the average, and insufficient weight where it was in defect. Consequently, the mortality table resulting from this process would on the whole overestimate the mortality throughout. In other words, the use of the unadjusted values of q introduces a systematic or "biassed" error into the calculations. If this is avoided, however, a very rough approximation to the graduated curve of q will give weights sufficiently near the truth for practical purposes, as a slight change in the relative weights of a given series of observations produces but little result upon the final solution.

THIRD LECTURE.

I PROPOSE in the present lecture to consider generally the use of frequency curves in relation to actuarial statistics. We have seen that the graphic method of dealing with these statistics, as also methods based upon finite difference formulæ, assume only that the true law of the series, if known, would be found to be represented by a continuous curve amenable to the ordinary processes of interpolation. It is often possible, however, to see that the ungraduated series can be well represented by a curve of a certain distinct character, and when this is found to be the case more satisfactory results are obtained, particularly where the data are few, by fitting to the original series a curve corresponding to its observed general character, so determining the constants in the equation of the curve as to secure the closest agreement with the ungraduated curve. If for example we turn to the series in column (2) of Table I, it will be at once seen that the general character of the series accords very closely to the "normal" frequency curve, or to some curve having the same general features. When we find that, by giving suitable values to the constants, a frequency curve can be made to fit the observations within the limits of the errors of observation we may be satisfied that the graduated curve thus produced is probably a better representation of the original than any that would result from a graphic or finite difference method of graduation.

Any curve which exhibits the law of variation in a particular function, such as a table of l_x, d_x or μ_x, may be considered for our purpose as a frequency curve. The expression is usually, however, confined to that class of curves which experience seems to show to be specially applicable to the observed distributions of deviations from mean values in statistical tables. We have already seen examples of such

tables where the frequency of the deviations of measures from their mean value follows certain comparatively simple laws. Professor Karl Pearson has examined a considerable variety of statistical data (mainly, but not entirely, biological) and finds that in practically all the cases examined the distribution of the various measurements may be represented fairly closely by one or other of the class of curves derived from the differential equation

$$\frac{1}{y}\cdot\frac{dy}{dx}=\frac{bx-x^2}{a-bx-cx^2}. \quad . \quad . \quad . \quad . \quad . \qquad (1)$$

where x represents the magnitude of a given deviation from the mean of a series of measures and y the frequency of such deviation.

As this group of curves is of considerable importance, though less so perhaps in relation to actuarial than in relation to some other classes of statistics, it is convenient to consider them first. It is not necessary here to discuss these curves analytically; the student may be referred to the original papers of Professor Karl Pearson*, or to an admirably condensed résumé by Mr. Robert Henderson in the *Journal of the Actuarial Society of America*, reprinted *J.I.A.*, xli, 429-442; and to Mr. W. Palin Elderton's treatise on "Frequency Curves and Correlation" in which Professor Pearson's methods are fully described. The table at the end of these lectures, which gives a sufficiently complete summary of such of the algebraical properties of these curves as are most useful in practice, is, with some unimportant modifications, based upon that given by Mr. Henderson in his paper. It will be sufficient for our present purpose to give a brief general description of these curves and of their use in connection with actuarial data.

We have already seen that the general character of curves, such as those of Tables I and II, is approximately determined by the average value of the squares and cubes of the deviations of the variable from its mean value; the former giving a measure of the compactness or diffuseness of the curve that is of the average extent of the deviations from the mean irrespective of their direction; the latter a measure of their departure from symmetry, or of the "skewness", of the curve. It will be useful at this point somewhat to extend

* *Phil. Trans.*, vol. 186, p. 343; vol. 197, p. 443, &c.

this general statement, and, before proceeding to a description of particular curves, to explain more in detail what is meant by the "moments" of a curve.

If we suppose $y = f(x)$ to represent the equation to a given curve, x varying between the limits h and k, the total area of the curve will be represented by the expression:

$$\text{area} = \int_k^h ydx.$$

We may suppose, for instance, to give definiteness to our ideas, that the function y represents the numbers under observation between age x and $x+dx$, the number of "years of life" observed between these ages being ydx, and the area of the curve, the sum of all these quantities, being the total years of life observed at all ages. If we now multiply each value of ydx by the corresponding age x and divide the total of these products by the total number of the "exposed", we shall have the average age of the whole. Put into symbols:

$$\int_k^h xydx \div \int_k^k ydx = \text{average value of } x. \quad . \quad . \quad . \quad (2)$$

$$= \text{1st moment of the curve round the ordinate for which } x = 0.$$

$$= m_1, \text{ say.}$$

Similarly,

$$\int_k^h x^n y\,.\,dx \div \int_k^k y\,.\,dx = \text{average value of } x^n$$

$$= n\text{th moment round ordinate for which } x = 0.$$

$$= m_n.$$

The moments of the curve may be taken round any ordinate we please. If, for example, the average value of x as found by equation (2), is x_1, then the ordinate corresponding to this value of x passes through the centre of gravity of the curve, and is termed the "centroid vertical." In general it is most convenient to take the value of the moments of the curve round this centroid vertical, for which obviously the first moment vanishes. The expression for the nth moment round this ordinate then becomes:

$$\int_k^h (x-x_1)^n ydx \div \int_k^h ydx = \mu_n \quad . \quad . \quad . \quad . \quad . \quad (3)$$

the average value of the nth power of the deviations $(x-x_1)$ between the values of x and the mean value. When the moments of a curve are spoken of without qualification, it will be understood that they are the moments round the "centroid vertical." These moments are, of course, those already referred to in Lecture I., p. 7, as representing the sums of the powers of the deviations of x from its mean value.

The following formulæ, which may be readily demonstrated,* connect the values of the moments round the "centroid vertical" with the moments round the ordinate for which $x=0$. Using the same notation as above, we have

$$\left.\begin{aligned}
&\mu_0=m_0=1\\
&\mu_1=0\\
&\mu_2=m_2-(m_1)^2\\
&\mu_3=m_3-3m_1m_2+2(m_1)^3\\
&\mu_4=m_4-4m_1m_3+6(m_1)^2m_2-3(m_1)^4
\end{aligned}\right\} \quad . \quad . \quad . \quad (4)$$

where the law of the coefficients is sufficiently obvious.

For the particular family of curves arising from the differential equation (1) formulæ may readily be found for the moments involving the various constants of the curves, and inversely, the values of the constants can be expressed in terms of the moments. The formulæ for the higher moments being sometimes complicated, it is more convenient to tabulate certain functions of the moments, *e.g.*:

$$\beta_1=\frac{(\mu_3)^2}{(\mu_2)^3};\quad \beta_2=\frac{\mu_4}{(\mu_2)^2};\quad \gamma=\frac{\beta_1+4}{\beta_2+3}$$

from which the constants of the curves may be obtained more readily, which are also useful in discriminating between the curves applicable to a given set of observations.

* *See* Elderton, p. 17–19; Henderson, *J.I.A.*, xli, 431–2.

The various curves arising from the differential equation (1) may, for our present purpose, be conveniently classified as under :—

Class I. Symmetrical curves. Range limited.
„ II. „ „ „ unlimited.
„ III. Skew curves. Range limited in both directions;
„ IV. Skew curves. Range limited in one direction;
„ V. Skew curves. Range unlimited in either direction;

the various types of curve being as follow. It will be seen that some of these Classes are repesented only by a single type of curve:

Class I. Symmetrical curves of limited range.—In this class we have only the single curve.

Type 1.
$$y=\kappa\left(1-\frac{x^2}{a^2}\right)^m$$

The values of x range from $+a$ to $-a$, for either of which values of the variable y becomes zero.

The average value of x is obviously zero, the corresponding ordinate y is a maximum, and clearly bisects the area enclosed between the curve and the axis of x. In other words, the "mean", "mode", and "median" of the curve all coincide, as in all symmetrical curves.

The second moment of the curve

$$=\mu_2=\frac{a^2}{2m+3}$$

and the "standard deviation"

$$=\frac{a}{\sqrt{2m+3}}.$$

The fourth moment

$$=\mu_4=\frac{3a^2}{2m+5}\mu_2.$$

The value of m will usually be positive when y equals zero at both limits. If $m>0<1$ the curve cuts the base-line at an angle. If m is negative the value of y becomes infinite at both limits, and m is always >-1.

This curve has a close relationship with the symmetrical point binomial curve, whose terms are proportional to the

terms in the expansion of $(\frac{1}{2}+\frac{1}{2})^n$, the general term of which may be written

$$y=\frac{\kappa}{\underline{\left|\frac{n}{2}+x\right.}\,\underline{\left|\frac{n}{2}-x\right.}}.$$

[It will, of course, be understood that the κ's in these formulæ, and in others, are not identical, but simply stand for some constant in each case, the numerical value of which is determined by the area of the curve.]

The binomial curve, however, can be conveniently used only to represent the definite points corresponding to integral values of $\frac{n}{2}\pm x$, whereas Type 1 represents a continuous curve (Note D, p. 122). The data with which an actuary has to deal are generally in the latter form, for example, the numbers living, the number of deaths, withdrawals, &c., between the ages x and $x+1$, and although usually the number of terms in the series is so considerable that the curve may be treated as a series of points, on the other hand, a binomial having so many terms will not generally be found a suitable curve to employ. In most instances where a series can be fairly represented by the symmetrical binomial, it can also be fairly represented by Type 1, with possibly some slight difference in range, as will be seen later.

There are other symmetrical curves of limited range, which are in the nature of frequency curves, but which do not belong to the family of curves derived from equation (1): such, *e.g.*, as the curve

$$y=\kappa e^{-\frac{m}{a^2-x^2}}$$

which, however, we need not discuss here.

Class II. Symmetrical curves of unlimited range.—In this class are two curves belonging to the family with which we are dealing.

Type 2. $$y=\kappa e^{-x^2/c^2}. \qquad (5)$$

This is the curve of "facility of error", or the "normal" frequency curve.

The average value of x is clearly zero, corresponding to the "mode" or the maximum value of y, and to the median.

The second moment $=\mu_2=\frac{c^2}{2}$, and the standard deviation$=\frac{c}{\sqrt{2}}$.

Type 1 evidently transposes into this curve when the value of a, and hence the range of the curve, is made indefinitely great. If we put $\frac{a^2}{m}=c^2$, making both a^2 and m indefinitely great, but their ratio finite, we have

$$\text{Limit}\left(1-\frac{x^2}{a^2}\right)^m = e^{-x^2/c^2} \quad \ldots \ldots \ldots \quad (6)$$

Even when the range of the curve is not great, that is when m and a^2 are not large numbers, there is a fairly close agreement between curves of Types 1 and 2 and the symmetrical binomial.

This may be seen by a numerical example, the following table showing

1. The values of $y=\frac{36{,}000}{\underline{|3+x}\,\underline{|3-x}}$ for integral values of x, these values being proportionate to the terms in the expansion of the binomial $\left(\frac{1}{2}+\frac{1}{2}\right)^6$.
2. The values of $y=993\left(1-\frac{x^2}{24}\right)^{6\cdot5}$.
3. The values of $y=1{,}026e^{-x^2/3\cdot1}$,

the constants in the two latter curves being chosen to give as good general agreement as practicable with the binomial curve.

Table VI.

Showing Similarity of Types 1 and 2 to the Symmetrical Point Binomial.

Values of Variable x	Binomial curve $y=\frac{36000}{\underline{\|3+x}\,\underline{\|3-x}}$	Type 1 $y=993\left(1-\frac{x^2}{24}\right)^{6\cdot5}$	Type 2 $y=1026e^{-x^2/3\cdot1}$
(1)	(2)	(3)	(4)
−4	0	2	6
−3	50	47	56
−2	300	303	282
−1	750	752	743
0	1,000	993	1,026
1	750	752	743
2	300	303	282
3	50	47	56
4	0	2	6
Totals ...	3,200	3,201	3,200

Had the range of the curves been greater, the binomial being taken to a higher power, and the values of the constants a^2 and m in col. (3) and of c^2 in col. (4) been larger, the agreement of the three curves would have been correspondingly closer. As it is, the two first curves are very nearly identical, while the "normal" curve, although theoretically of unlimited range, is fairly close to the binomial, the terms corresponding to values of x numerically greater than 4, amounting to less than 1 in the aggregate. It will be noticed that the values of y in the limited curves necessarily diminish more rapidly as the limiting values of x are approached, while the normal curve is less flat in the centre.

Type 3. $$y=\kappa\left(1+\frac{x^2}{a^2}\right)^{-m} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

This curve, which is also symmetrical and unlimited in range, diverges from the normal curve in a direction opposite to Type 1, the values of y diminishing, when x is large, more slowly than in the normal curve. The curve transposes into the latter (Type 2) when a^2 and m are indefinitely large, $\frac{m}{a^2}=c^2$ being, however, finite. We then have

$$\text{Lt. } \kappa\left(1+\frac{x^2}{a^2}\right)^{-m}=\kappa e^{-x^2/c^2}.$$

The average value of x in the curve $y=\kappa(a^2+x^2)^{-m}$ is zero, corresponding again to the "mode"; the second moment $=\mu_2=\frac{a^2}{2m-3}$ and the "standard deviation" $=\frac{a}{\sqrt{2m-3}}$. The fourth moment $=\mu_4=\frac{3a^2}{2m-5}\mu_2$ and, it is clear, becomes infinite unless $m>\frac{5}{2}$. Indeed, the higher moments of the curve must become infinite whatever be the value of m.

The classes of symmetrical curves are of somewhat limited application to actuarial statistics, although there are certain cases in which they represent the observations fairly well.

Class III. Skew curves. Range limited in both directions.—There is only a single curve of this class in the family of curves we are considering, namely:

Type 4. $$y=\kappa\left(1-\frac{x}{a}\right)^{m_1}\left(1+\frac{x}{a}\right)^{m_2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

The values of x range from $-a$ to $+a$; the "mode" is at $x=\frac{m_2-m_1}{m_1+m_2}\cdot a$, for which value y is a maximum; the mean value of x is $\frac{m_2-m_1}{m_1+m_2+2}\cdot a$. The expressions for the moments of the curve are simplified by putting it into the form given in the table on p. 140. If we write $m_1=np-1$, and $m_2=nq-1$ (where $p+q=1$), the equation to the curve (which does not, of course, change in character with this transposition) becomes

$$y=\kappa\left(1-\frac{x}{a}\right)^{np-1}\left(1+\frac{x}{a}\right)^{nq-1} \quad . \quad . \quad . \quad . \quad . \quad (9)$$

the variable having the same range of values $-a$ to $+a$, the "mode" being at $x=\frac{n}{n-2}(q-p)a$; the average value of $x=(q-p)a$; the second moment $=\mu_2=\frac{4pq}{n+1}\cdot a$, and the "standard deviation" the square root of this quantity.

When $m_1=m_2=m$, this Type evidently transposes into Type 1, and thence into Type 2 when m is infinite.

This curve is related to the skew point binomial arising from the expansion of $(p+q)^{n'}$, where p and q have approximately the same values as in equation (9), and where the index of the binomial is not too small, there is a fair numerical agreement, as may be seen in the following table, where the figures given in col. (2) are proportional to the terms in the binomial expansion of $\left(\frac{1}{3}+\frac{2}{3}\right)^6$:—

Table VII.

Showing Numerical Similarity of the Curve of Type 4 with the Skew Binomial.

Value of Variable x	Binomial curve $y=\frac{5760}{\lfloor 3+x \lfloor 3-x}\cdot 2^x$	Type 4 $y=\kappa(4{\cdot}75-x)^{5{\cdot}51}(6{\cdot}25+x)^{11{\cdot}56}$
(1)	(2)	(3)
-4	0	0
-3	1	1
-2	12	13
-1	60	61
0	160	159
1	240	240
2	192	194
3	64	60
4	0	1
Totals ...	729	729

It will be seen that for so small a value of n as 6 the binomial curve can be closely represented by means of selected points in the continuous curve of Type 4. When the value of n is large, a much closer agreement is obtainable.

The skew binomial is of importance to the actuary as representing the law of the deviations between the actual number of events observed in a given series of trials and the "expected" number when computed by the true value of the probabilities. There are very many statistical distributions capable of being well represented by the binomial curve if the latter is treated as a continuous curve. This procedure is not, however, convenient in practice, as it rarely happens that the given ordinates coincide with the integral values of x in the general term $\frac{\underline{|n}}{\underline{|x}\,\underline{|n-x}}\left(\frac{p}{q}\right)^x$, and, moreover, the analysis, when the curve is treated as continuous, is not very simple. (*See* Note D, p. 122.)

The form of curve corresponding to Type 4 varies very considerably with certain changes in the values of the constants m_1 and m_2. In its more usual form, when both m_1 and m_2 are >1, as in Table VII, the curve bears a general resemblance to the age distribution of the "entrants" in a mortality, or similar experience (*see* Table II), also to the numbers of the exposed to risk; to the number of marriages, or to the rate of marriage at various ages; to the average number of children under age, or to the cost of their pensions at the death of the father, a function of use in pension fund valuations; to the number of retirements in such funds where superannuation occurs on invalidity and not at a specified age; to the incidence of attacks, or of mortality, from certain diseases, &c. Owing to the number of constants involved (as the increment of x may represent any period of time, there are virtually five), the curve is very adaptable.

It will be readily seen that if the values of both m_1 and m_2 in equation (8) are high the curve makes very close contact with the axis of x at either limit; if m_1 or m_2 lies between 0 and 1, the curve meets the axis of x at an angle; whereas, if either or both of them are negative, the expression becomes infinite at one or both limits. The area of the curve and the moments do not, however, become infinite if both m_1 and m_2 are greater than -1.

Class IV. Skew curves. Range limited in one direction.—There are two curves of this class.

Type 5. $$y=\kappa x^m e^{-x/a} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

which is a limiting form of curve No. 4, the values of x ranging from 0 and ∞.

The "mode" is at $x=ma$; the mean value of x is $(m+1)a$; the second moment $(m+1)a^2$; and the third moment $2(m+1)a^3$; these being sufficient to determine the constants.

In the usual form of the curve, that is when $m>1$, this curve represents fairly well some of the statistical distributions represented by curve No. 4. Owing to the feature that as x becomes large the successive terms have a tendency to run into a geometrical progression, it is not so well suited to such distributions as that of the "exposed to risk", where the effect of the rapid rise in the rate of mortality at the older ages makes itself felt in an increasingly rapid diminution in the values of y. This is somewhat unfortunate, as the curve is a simple one, determined by the values of its first three moments, and except for the reason stated, well suited for use in connection with Makeham's formula for the force of mortality.

As in Type 4, the character of this curve may be entirely changed by an alteration in the values of the constant m. If this constant vanishes the curve becomes a diminishing geometrical progression; while for negative values of m the curve becomes infinite at the lower limiting value of x. The value of m must in any case >-1.

The actuary has to deal with several distributions roughly similar to a diminishing geometrical progression as, for example, the curve of infant mortality, the rate of withdrawal in successive policy years, or the difference between the select and ultimate mortality rates in a select mortality table. Other expressions giving a similar form of curve may be employed to represent these distributions as, for example, $y=\kappa\,(a+e^{-mx})$, with a minimum value of κa when x is very large; or $y=\kappa(x+a)^{-m}$, where if a is small we have a curve again similar to that of infant mortality, x representing the age.

Type 6. $$y=\kappa\left(\frac{x}{a}-1\right)^{m_1}\left(\frac{x}{a}+1\right)^{-m_2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

where the limiting values of x are a and ∞, with an

average value of $x=\frac{m_2+m_1}{m_2-m_1-2}.a$; the "mode" occurring at $x=\frac{m_2+m_1}{m_2-m_1}a$. The expressions for the moments are much simplified by writing the equation to the curve in the form given in the Table on pp. 140–1.

Type 7. $$y=\kappa x^{-m}e^{-a/x}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

Where x varies between 0 and ∞, having an average value of $\frac{a}{m-2}$, with the "mode" at $x=\frac{a}{m}$. The second moment $\mu_2=\frac{a^2}{(m-2)^2(m-3)}$ and the "standard deviation" consequently $=\frac{a}{(m-2)\sqrt{m-3}}$.

Here m must be >3, or the second moment becomes ∞, and the fourth moment becomes infinite unless m is greater than 5.

Neither this nor the preceding curve are of any wide application in actuarial statistics, owing to the fact that the values of y for large values of x diminish with increasing slowness; a feature not often met with in practice except in such a function as the "rate of withdrawal." The same remark holds good of the single curve constituting Class V.

Class V. Skew curves. Range unlimited in either direction.

Type 8. $$y=\kappa\left(1+\frac{x^2}{a^2}\right)^{-m}e^{\nu\tan^{-1}\frac{x}{a}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

This is the only skew curve of this family having unlimited range. The average value of $x=\frac{\nu}{2(m-1)}a$; the "mode" is at $\frac{\nu}{2m}a$.

The expressions for the moments and their functions are simplified by writing $\left(\frac{n}{2}+1\right)$ for m in equation (13), as in the Table on pp. 140–1. For the reason stated above, the curve is not specially useful to Actuaries.

Assuming that a given statistical series can be represented by one or other of the curves above described, the appropriate curve can be found by means of certain criteria based upon an examination of the "moments" of the curve; that is to say, the sums of the powers of the deviations from the mean value. These criteria are furnished by the table on pp. 140–1, above referred to.

As the calculation of the criterion is somewhat lengthy, it may be noted that if the logarithms of y are tabulated for equal intervals of the variable x, and the values of $\Delta^2 \log y$ taken out, these give us information as to the nature of the curve. The value of $\Delta^2 \log y$ will be constant and negative for the "normal" curve Type 2; negative and symmetrical with a minimum numerical value in the centre of the range, for Type 1, or for any binomial curve; uniformly negative, non-symmetrical, and with a numerical minimum in the case of Type 4 (where this curve vanishes at the limits); and uniformly negative and continuously decreasing towards the upper limit of x in the case of Type 5, where this curve vanishes at the limits.

In the case, therefore, of those curves most useful to the Actuary the function $\Delta^2 \log y$, computed for the ungraduated curve, enables us to select generally the formula most suited to the series. For this purpose if the data are grouped it will generally be better to compute the approximate values of the central ordinates of each group by an interpolation formula, such as that given on p. 57.

Other types of curves will sometimes be found useful besides those arising from the differential equation on p. 39; but they do not generally lend themselves so readily to the method of moments.

If, for example, we write

$$y=\kappa\cdot e^{-\left(\frac{m}{a+x}+\frac{n}{b+x}\right)} \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

we obtain, when m and n are numerically unequal, a skew curve vanishing when $x=-a$ or $-b$. We may deal with this curve in practice by determining the values of equidistant ordinates as shown on pp. 57–8. Thus

$$\log y=\kappa'-\frac{m}{a+x}-\frac{n}{b+x}=w \quad . \quad . \quad . \quad . \quad (15)$$

As $\log y$ becomes $-\infty$ at the limits, we multiply both sides by $(a+x)(b+x)$, thence

$$\begin{aligned} w[ab+(a+b)x+x^2] & \\ &= \kappa'[ab+(a+b)x+x^2]-m(b+x)-n(a+x) \\ &= \mathrm{A}+\mathrm{B}x+\mathrm{C}x^2 \text{ (say)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16) \end{aligned}$$

where the unknowns are a, b, A, B and C.

If we difference three times the right hand side vanishes and we have a series of expressions involving (ab) and $(a+b)$ equated to zero and by suitably grouping these, or by using the method of moments a and b, and thence the remaining constants, may be evaluated.

A similar process may be employed with advantage with a curve such as the usual form of exposed to risk or died, when the data are in large age groups. We may then take w in equation (15) to represent the common log of the ratio of the numbers above age x to the numbers below age x in the series. That is, if the total number in the series $=\mathrm{N}$, the number above age $x=\mathrm{Y}$, we may write

$$\log\left(\frac{\mathrm{Y}}{\mathrm{N}-\mathrm{Y}}\right)=w=\mathrm{K}'-\frac{m}{a+x}-\frac{n}{b+x} \quad . \quad . \quad . \quad (17)$$

In many cases the constant K′ may be omitted if the number of groups is small; in this case C in equation (16) becomes zero. On the other hand it may sometimes be found necessary to add a term to the right hand of equation (16) involving x^3.

FOURTH LECTURE.

We shall now consider very shortly the problem of fitting frequency curves to statistical data. To do this at length would be impossible in the time at our disposal, and the student who wishes to pursue the subject in detail may read the original papers, already referred to (p. 39), of Professor Karl Pearson, to whom the development of the subject is due, or Mr. Elderton's book. There are certain general principles however, which may be usefully considered. The method usually employed in fitting these curves is by making the moments of the graduated equal to those of the ungraduated curve, which is equivalent to making the quantities Σ (deviations), Σ^2 (deviations), &c., as far as Σ^4 or Σ^5 equal to zero. This method may not always be the most convenient or the best for the purpose of the Actuary, but it is so for most statistical purposes, and has come much into use accordingly.

We have already seen that, in the case of the curves arising from the differential equation on p. 39, expressions for the moments may be obtained in terms of the constants which will enable us to determine the value of the constants, when the numerical value of the moments is known. For the purpose of fitting the appropriate curve to any given series of observations it is only necessary to determine the value of the moments as given by the observations, that is, the value of the sum of the squares, cubes, &c., of the deviations from the mean value of the variable.

It will be useful to consider shortly the calculation of the numerical value of the moments in a given instance. Take first the simplest possible case where we have to do not with a continuous curve, but with a series of points representing isolated ordinates, where in consequence we replace integra-

tions by summations. In the following table, the first column contains the values of the independent variable x, the range of values being from 0 to 6. The second column contains the values of its function y, which are proportionate to the successive terms in the expansion of the binomial $\left(\frac{1}{3}+\frac{2}{3}\right)^6$, the constant multiplier 729 being introduced merely to avoid fractions. The remaining columns, in which the average value of x and the values of the successive moments are worked out, explain themselves. It may be remarked that in this example the average value of x, and the deviations from the average, are all integral, and it is therefore convenient to calculate at once the moments round the average value ("centroid vertical"). In most cases, however, the average and the deviations will not be integral, and then it will be more convenient to calculate the moments round the origin or some selected middle value of the variable, afterwards transferring the moments to the mean by the formulæ given on p. 41.

TABLE VIII.

Moments of the Point Binomial Curve.

$$729 \cdot \frac{\underline{|6}}{\underline{|x}\,\underline{|6-x}}\left(\frac{2}{3}\right)^x\left(\frac{1}{3}\right)^{6-x} = \frac{720}{\underline{|x}\,\underline{|6-x}}\cdot(2)^x.$$

x	y	xy	$(x-4)y$	$(x-4)^2y$	$(x-4)^3y$	$(x-4)^4y$
0	1	0	-4	16	-64	256
1	12	12	-36	108	-324	972
2	60	120	-120	240	-480	960
3	160	480	-160	160	-190	160
4	240	960	0	0	0	0
5	192	960	192	192	$+192$	192
6	64	384	128	256	$+512$	1,024
Totals	729	2,916	0	972	-324	3,564
Totals $\div 729$	1	4 mean value of x	0 $=\mu_1$	$\frac{4}{3}$ $=\mu_2$	$-\frac{4}{9}$ $=\mu_3$	$\frac{44}{9}$ $=\mu_4$

Obviously, when the moments are calculated about the mean the first moment is zero (because it represents the average

deviation from the average value). The even moments are always positive, because each term is of the form $y_x x^{2n}$, *i.e.*, essentially positive; and if the curve is symmetrical the odd moments vanish, because each term of the form $y_x x^{2n+1}$ is cancelled by a term (equidistant from the mean) of the form $y_{-x}(-x)^{2n+1}$. In general, where the curve is not symmetrical, the third, fifth, &c., moments will not be zero.

In the above illustration, we have considered x to have integral values only. This may be said to approximate to the conditions of many statistical tables used by the Actuary where x represents the year of age under observation, and where it is indifferent whether the observations are supposed to be spread over the year in the form of a continuous curve, or whether we consider them all to have reference to the central point of the year. In these cases, however, x will generally have a large range of values, amounting possibly to 60 or 80, and the labour of computing the numerical value of the moments is then much lessened by grouping the facts in larger sections, though we cannot then safely assume the totals of each group to be concentrated at the middle ordinate.

Take the set of observations in Table IX representing for decennial age groups numbers exposed to risk in the middle of each year of age, *i.e.*, $\overline{E}_x = E_x - \frac{1}{2}\theta_x$, in the recent mortality experience of lives assured by ascending premium policies,* excluding the first ten years from entry. Here we have no longer the values of equidistant ordinates of the curve, but the area of the curve enclosed between successive ordinates. To obtain the moments of the curve with any degree of accuracy, we cannot treat these areas as proportional to their central ordinate.

It will be noticed that the particular curve we are dealing with becomes gradually zero at either extremity,† and we may assume, without serious error, that it makes "close contact" at either end with the axis of x, that is to say, is asymptotic thereto. In these cases, Mr. Sheppard has shown‡ that very approximate values for the moments may be found

* *See* Unadjusted Data, Minor Classes of Assurances, p. 191.

† We omit the numbers at risk under age 25 (arising from entrants under age 15), amounting to only 25 in all.

‡ An elementary demonstration is given in Elderton's Treatise, p. 28–29.

by treating the area of each successive section of the curve as concentrated in the middle ordinate of the section; in other words, treating the values of y as representing isolated ordinates exactly as was done in Table VIII; and then applying to the values of the moments so found (denoted by the symbol m') the following adjustments leading to the corrected moments denoted by the symbol m:—

$$m_1 = m'_1$$

$$m_2 = m'_2 - \frac{1}{12}$$

$$m_3 = m'_3 - \frac{1}{4}m' = m'_3 - \frac{1}{4}m_1$$

$$m_4 = m'_4 - \frac{1}{2}m'_2 + \frac{7}{240} = m'_4 - \frac{1}{2}m_2 - \frac{1}{80}.$$

For moments round the centroid vertical these become, remembering that $\mu_1 = 0$,

$$\mu_2 = \mu'_2 - \frac{1}{12}$$

$$\mu_3 = \mu'_3$$

$$\mu_4 = \mu'_4 - \frac{1}{2}\mu_2 - \frac{1}{80}.$$

TABLE IX.

Ascending Premium Assurances—Experience 1863–1893.

Duration 10 *years and upwards.*

Calculation of Moments of "Exposed to Risk" Curve.

Ages	Exposed to Risk y	x	xy	x^2y	x^3y	x^4y
25–35	2,874	−2	− 5,748	11,496	−22,992	45,984
35–45	22,020	−1	−22,020	22,020	−22,020	22,020
45–55	26,164	0	...	...	...	...
55–65	17,391	1	17,391	17,391	17,391	17,391
65–75	7,845	2	15,690	31,380	62,760	125,520
75–85	1,761	3	5,283	15,849	47,547	142,641
85–95	81	4	324	1,296	5,184	20,736
Totals	78,136	...	10,920	99,432	87,870	374,292
Reduced to unit area	1	...	·13976 $=m'_1$	1·2725 $=m'_2$	1·1246 $=m'_3$	4·7903 $=m'_4$

From these results we obtain by means of the corrections above stated—

$$m_1 = \cdot 13976\,;\quad m_2 = 1\cdot 1892\,;\quad m_3 = 1\cdot 0897\,;\quad m_4 = 4\cdot 1832.$$

Whence, by the equations on p. 41.

$$\mu_2 = 1\cdot 1697\,;\quad \mu_3 = \cdot 5965\,;\quad \mu_4 = 3\cdot 7122.$$

If quinquennial age groups had been used, making due allowance for the unit of time still being taken as ten years, the corresponding values would have been

$$m_1 = \cdot 13848\,;\quad \mu_2 = 1\cdot 1741\,;\quad \mu_3 = \cdot 5869\,;\quad \mu_4 = 3\cdot 7160.$$

using these latter values, as the more accurate, we obtain for the values of the functions β_1, β_2 and γ.

$$\beta_1 = \mu^2_3/\mu^3_2 = \cdot 21283\,;\quad \beta_2 = \mu_4/\mu^2_2 = 2\cdot 6957\,;\quad \gamma = \frac{\beta_1 + 4}{\beta_2 + 3} = \cdot 7397\,;$$

As μ_3 does not vanish, and γ is $> \frac{2}{3}$, we see from the table on pp. 140–1, that if the series can be represented by any of the curves there given, it must be by No. 4, excluding the skew binomial as unsuitable for reasons already given. It is also obvious from the run of the figures in Table IX, that the curve is limited in both directions. Equating the expressions in Table IX with the above numerical values, we have

$$\gamma = \frac{2}{3}\cdot\frac{n+3}{n+2} = \cdot 7397\,;\ \text{whence } n = 7\cdot 13$$

$$\beta_1 = \frac{4(n+1)(p-q)^2}{(n+2)^2 pq} = \cdot 21283\,;$$

whence

$$(p-q)^2 = \cdot 5453pq$$

$$(p+q)^2 = 4\cdot 5453pq = 1 \ (\text{since } p+q=1)$$

$$(p-q) = \sqrt{\frac{\cdot 5453}{4\cdot 5453}} = \cdot 3464$$

giving

$$p = \cdot 6732\,;\quad q = \cdot 3268$$

$$\mu_2 = \frac{4pq}{n+1}\cdot a^2 = 1\cdot 1741\,;\ \text{whence } a = 3\cdot 293$$

thus giving a range of 32·93 years on either side of the age for which the value of x in the formula$=0$. This has nothing to do with the zero point (age 50) in Table IX. The mean age as is seen from that table is $50+1\cdot 385=51\cdot 385$. The value of m, the mean as computed by the above formula, is

$$m_1 = (q-p)a = -1\cdot 1407$$

that is, 11·407 years earlier than the central point of the range, giving for the latter, $51·385+11·407=62·79$, say. The range of the curve is therefore from age 29·86 to age 95·72; and, computing the values of $np-1$ and $nq-1$, we have, for the final form of the equation of the curve, when x=the age:

$$y=\kappa.(x-29·86)^{1·33}(95·72-x)^{3·80}$$

It is often a convenience, however, to have the values of the central ordinates of the groups, which may be approximately obtained by interpolation. If the numbers in any group are represented by the symbol u_x, the number of years in each group being t, the value of the central ordinate of the group (that is to say, the numbers under observation exactly at the central age of the group) will be approximately $\frac{1}{t}\left(u_x-\frac{\Delta^2 u_{x-1}}{24}\right)$*. As, however, it is convenient to treat the interval t as the unit, for the time being, we may write as the values of the central ordinates $u_x-\frac{\Delta^2 u_{x-1}}{24}$ (the original numbers for each group less $\frac{1}{24}$th of their respective central second differences). In the class of curves we are discussing, namely, those having close contact at both ends with the axis of x, the numerical values of the moments as deduced from these ordinates will be very nearly the values for the continuous curve, unless the number of groups is very small. Thus the values of $\int_k^h ydx$, and of the functions $\int_k^h xydx$, $\int_k^h x^n ydx$, will be found by taking the *sum* of the ordinates of y, computed as above, and the *sum* of the products xy, $x^n y$.

An advantage attaching to the use of ordinates in lieu of areas is that, in the class of curves we are dealing with, we can, by examination of the differences of the logarithms of the ordinates, gain a better idea of the nature of the curve than can be obtained from the grouped figures. (*See* Third Lecture, p. 50.) It is also easier to compare the graduated figures as given by the frequency curve by means of isolated ordinates than by means of groups or areas.

* The formula to 4th differences is $u_x-\frac{\Delta^2 u_{x-1}}{24}+\frac{\Delta^4 u_{x-2}}{213}$ nearly, and in order that the resulting 4th moment should agree exactly with that obtained from the use of the grouped figures, or areas, with Sheppard's corrections, the 4th difference is required, but for practical purposes it is not often needed.

The use of the central ordinates of the groups has the incidental advantage, which is very considerable in the case of a mortality or similar experience, of giving trustworthy values of the force of mortality, or corresponding function, for the ages corresponding to the position of the ordinates. In the usual plan of summarising a mortality table by giving the numbers at risk and deaths in consecutive age groups, the ratio of the deaths to the numbers at risk in each group is not a useful function, as it does not correctly represent the mortality for the central age, except near the middle of the table, where the numbers under observation in successive years is nearly constant.

We may apply this method to the example already dealt with on p. 55, viz., the experience of ascending premium policies. The calculations as set out in the following tabular form are sufficiently clear:

TABLE X.

Mortality experience of lives assured by ascending Premiums, 1863–1893. *Duration* 10 *years and upwards.*

Ages	Central age of group (x)	Exposed to Risk	Died	*ESTIMATED CENTRAL ORDINATES		μ_x Central Age
				Exposed to Risk	Died	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
25–30	27·5	266	2	168	·8	·0048
30–35	32·5	2,607·5	31	2,448	29·2	·0119
35–40	37·5	8,788	102	8,860	102·0	·0115
40–45	42·5	13,232·5	173	13,389	175·2	·0131
45–50	47·5	13,910	192	14,007	191·7	·0137
50–55	52·5	12,254	218	12,284	218·6	·0178
55–60	57·5	9,878·5	229	9,878	228·4	·0232
60–65	62·5	7,512·5	255	7,518	255·4	·0340
65–70	67·5	5,007·5	271	4,994	274·4	·0549
70–75	72·5	2,837	206	2,809	205·6	·0732
75–80	77·5	1,347·5	151	1,324	151·4	·1144
80–85	82·5	413·5	85	389	84·8	·2180
85–90	87·5	77	24	66	22·3	·3379
90–95	92·5	4·5	3	2	2·2	1·1000
Totals...	...	78,136	1,942	78,136	1,942·0	...

Taking 5 years as the unit, computing by formula $u_x - \frac{\Delta^2 u_{x-1}}{24}$, where u_x represents the number in columns (3) and (4). By this formula there are −11 persons exposed to risk at age 22·5; these have been included in the group 25–30.

If the values of the moments are computed from columns exactly as was done with the Binomial Curve (Table VIII, p. 53) they will be found to be practically identical with those found above. The estimated values of μ_x for the central ages of the group are inserted as they will be used later.

In many cases the principle of the method of moments may be used to fit a curve to a series of observations without actually computing the numerical values of the moments themselves, using instead the successive summations of the ordinates, or areas, from which the moments can be readily obtained if required. This method is also useful if one or both limits to the range of the curve can be assumed.

Consider a scheme such as the following, in which, with a view to clearness, we use actual numbers of the series, given on p. 53, instead of symbols:—

x	u_x	Σu_x	$\Sigma^2 u_x$	$\Sigma^3 u_x$	$\Sigma^4 u_x$	$\Sigma^5 u_x$	$u_x \times x^4$
0	1	729	...	...	...	...	0
1	12	728	2,916	7,776	...	...	12
2	60	716	2,188	(6,318) 4,860	9,180	15,660	960
3	160	656	1,472	2,672	4,320	(11,070) 6,480	12,960
4	240	496	816	1,200	1,648	2,160	61,440
5	192	256	320	384	448	512	120,000
6	64	64	64	64	64	64	82,944
							278,316

In this scheme, each column is formed from the preceding by successive addition from the bottom, in the same way that the M_x column is formed from C_x, and R_x from M_x.

If we take the value against $x=0$ in the column Σu_x, say Σu_0, we see that each value of u_x occurs once only in that total. In the total appearing against $x=1$ in the second summation, say $\Sigma^2 u_1$, each value of u_x occurs x times; similarly the total against $x=2$ in the column $\Sigma^3 u_x$, say $\Sigma^3 u_2$, represents the sum of the products $\frac{x(x-1)}{2} u_x$; and the

total against $x=3$ in the column $\Sigma^4 u_x$, say $\Sigma^4 u_3$, represents the total of the products $\frac{x(x-1)(x-2)}{6} u_x$, and so on, the coefficients following the Binomial law. It is evident from this that the sums of the products $x^2 u_x$, $x^3 u_x$, &c., are implicitly contained in these totals; and that if these sums of the graduated and ungraduated values are in agreement, the moments of the two curves will also agree. Writing m_n as the value of the nth moment round the ordinate of $x=0$, we shall find:*

$$m_1 = \frac{\Sigma^2 u_1}{\Sigma u_0}$$

$$m_2 = \frac{2\Sigma^3 u_2 + \Sigma^2 u_1}{\Sigma u_0}$$

$$m_3 = \frac{6\Sigma^4 u_3 + 6\Sigma^3 u_2 + \Sigma^2 u_1}{\Sigma u_0}$$

$$m_4 = \frac{24\Sigma^5 u_4 + 36\Sigma^4 u_3 + 14\Sigma^3 u_2 + \Sigma^2 u_1}{\Sigma u_0}$$

These formulæ may be simplified if we write them in a form analogous to central difference formulæ—writing, for example:

$$\Sigma^3 u_{x+\frac{1}{2}} \text{ for } \frac{\Sigma^3(u_{x+1}+u_x)}{2},$$

these average values being shown in antique type in the Scheme. We then have, omitting the common divisor Σu_0:

$$m_1 = \Sigma^2 u_1$$

$$m_2 = 2\Sigma^3 u_{1\frac{1}{2}}$$

$$m_3 = 6\Sigma^4 u_2 + m_1$$

$$m_4 = 24\Sigma^5 u_{2\frac{1}{2}} + m_2$$

The equivalence of the above formulæ may be illustrated by the following numerical examples based on the above scheme.

* See the demonstration in Note E, p. 124.

Using N as an abbreviation of Σu_0=the total number of observations, we have

$$N.m_0= \quad 729=\Sigma u_0$$

$$N.m_1= \quad 2916=\Sigma^2 u_1$$

$$N.m_2= 12636=2\Sigma^3 u_2 +\Sigma^2 u_1 \quad =2\times 4860+2916$$
$$=2\Sigma^3 u_{1\frac{1}{2}} \quad =2\times 6318$$

$$N.m_3= 57996=6\Sigma^4 u_3 +6\Sigma^3 u_2 +\Sigma^2 u_1=6\times 4320+6\times 4860+2916$$
$$=6\Sigma^4 u_2 \quad +\Sigma^2 u_1=6\times 9180 \quad +2916$$

$$N.m_4=278316=24\Sigma^5 u_4+36\Sigma^4 u_3+14\Sigma^3 u_2+\Sigma^2 u_1=24\times 2160$$
$$+36\times 4320+14\times 4860+2916$$

$$=24\Sigma^5 u_{2\frac{1}{2}} \quad +2\Sigma^3{}_{1\frac{1}{2}} \quad =24\times 11070+2\times 6318$$

The last may be compared with the direct calculation of $x^4 u_x$ given in the last column of the scheme. The values of the moments through the centroid vertical may be obtained if required by the formulæ:

$$\mu_1=0$$

$$\mu_2=m_2-(m_1)^2$$

$$\mu_3=m_3-3(m_1)\mu_2-(m_1)^3$$

$$\mu_4=m_4-4(m_1)\mu_3-6(m_1)^2\mu_2-(m_1)^4.$$

Where the number of terms in the series is few, there is no special advantage in this method; but if the number of terms is considerable it effects a saving of time, more particularly if the calculation of the moments round the centroid vertical is not needed by the conditions of the problem, as in the case of the graduation of rates of mortality by Makeham's or any similar frequency formula.

The case of curves not making close contact with the axis of x at both ends requires to be considered separately, but the results obtained are not altogether satisfactory, *see* Elderton, pages 29-30. The difficulty can, however, to a great extent be avoided in most cases arising in actuarial work by using very small groups, or even individual values for each year of

age, &c., in calculating the moments. The labour although increased is by no means prohibitive if the summation method, above described, be adopted.

Professor Karl Pearson has shown* that the method of fitting a curve by computing its moments should lead to nearly the same results as the method of least squares. If we are fitting to a given set of observations an ordinary parabolic curve, represented by the equation $y=a+bx+cx^2+$&c., then the method of moments and the method of least squares are identical.† He infers from this fact that, even if y is represented by a more complex expression, the numerical results from the method will be nearly the same as with the method of least squares. It would appear at first sight that the effect of the method of moments is to give equal weight to each observation or group of observations, in spite of their having unequal average errors; whereas the method of least squares should, strictly speaking, be applied only when the average error of each observation is nearly equal.‡ In a mortality table, where the number of persons under observation and the number of deaths are relatively large in the middle of the table and fall off to zero at the beginning and end, the probability of a given error in the value of q is very much smaller at the central ages; while, on the other hand, the probability of a deviation of a unit in the number of deaths is correspondingly greater. The same applies to most tables of statistics, as they usually present a series starting from zero, rising to a maximum, and diminishing to zero again, the weight of the observations being in the middle of the curve, where, however, the probability of a given numerical deviation in the actual numbers is also greater.

We have seen that in a series of numbers representing the distribution of a group into sub-groups the average error in any given case is approximately $\cdot 8\sqrt{\frac{m(n-m)}{n}}$, where n is the number in the group and m the (graduated) number in the sub-group. If, as is generally the case, n is large compared to m,

* *Biometrika*, vol. i, p. 266–271.

† This assumes that the unadjusted moments (m not m') are used, *i.e.*, that the numbers represent ordinates and not areas. If the moments are assumed to represent areas and the corresponding corrections are introduced, the method of moments no longer gives precisely the same results as the method of least squares: *see* examples given by Todhunter, *J.I.A.*, xli, 444.

‡ *See* Note C, p. 117.

this expression may be taken as equal to $\cdot 8\sqrt{m}$, the average error in the ratio $\frac{m}{n}$ being approximately $\cdot 8\frac{\sqrt{m}}{n}$. Thus, if the number at risk at a given age equals n and the true probabilities of death and survivorship, are q and p, then $\cdot 8\sqrt{npq}$* (which as p is nearly unity for the greater number of ages may be roughly taken as $\cdot 8\sqrt{\text{number of deaths}}$), is an approximate expression for the average deviation from the expected number of deaths. The method of moments, if employed to represent a given series by a parabolic curve, assumes an equal probability of unit error in each term of the series. If, therefore, the series is of such a character that the extreme values are relatively small, these parts of the data will have somewhat less than their due weight in the fitting process. If, however, the formula to be fitted does not represent a parabolic curve, but a curve analogous to the normal curve ke^{-x^2/c^2}, say a curve of the form $e^{a+bx+cx^2+\&c.}$ then it will be found that, on the assumption that the mean error in any value y is equal to $\sqrt{y_1}$ (where y_1 represents the graduated value of y) the method of moments gives the same result as the method of least squares when the observations are duly weighted (*see* Note F, p. 129).

We come now to the class of curves representing not the actual numbers in statistical tables, but the ratios of the corresponding numbers in the double series, such as those of tables of "Exposed to Risk" and "Died", curves, that is, representing such functions as rates of mortality, of marriage, of lapse, of superannuation, &c. The most interesting and important of these is the curve due to Makeham's development of Gompertz's hypothesis, in which the force of mortality at a given age x is represented by the expression

$$\mu_x = A + Bc^x$$

leading to the equation

$$\log_{10} l_x = K + A'x + B'c^x.$$

This curve has a double value as, apart from its use in graduating a mortality table, it has the valuable property

* *See* Note A, p. 110; *J.I.A.*, xxvii, 214.

that the values of annuities on n joint lives of various ages can be found from a table of single entry showing the values of annuities on n lives of equal age. Owing to its importance it will be useful to give some attention to the problem of fitting this curve to a mortality experience. We will first consider the case of an aggregate or non-select table, that is, a table in which the rate of mortality is a function of the age alone.

Various methods have been employed to obtain the values of the constants A, B, c, corresponding to a given experience. That used by Makeham, and subsequently in a modified form by Woolhouse, is based on selected values of $\log l_x$ taken from a table already graduated by a finite difference formula. Four values of $\log l_x$ may be taken, covering practically the whole of adult life, say the values at ages 20, 40, 60, and 80, or 25, 45, 65, 85. Either set are sufficient to determine the four constants, K, A′, B′ and c, as above. In Woolhouse's graduation of the H$^{\text{M}}$ Table, both of these sets of ages were employed, the most advantageous values of the constants being found by comparing the deviations between the graduated and ungraduated values of l_x at quinquennial ages according to the two preliminary graduations. If a single set of four values of l_x is taken as the basis of the graduation, the effect is the same as employing the sums of the forces of mortality $(\mu_{x+\frac{1}{2}})$ between the selected ages, giving equal weight to the values at each age.

The method employed by Mr. King in the Institute of Actuaries' Text-Book, Part II., substitutes for graduated values of $\log l_x$ at isolated ages, the sum of certain groups of the ungraduated values of $\log l_x$. The effect of this method would appear to be to give a diminishing weight to the values of μ_x for the ages at the commencement and end of the table, which is so far in accordance with theory, and to eliminate the effect of errors in isolated values of l_x. In *Biometrika* (vol. i., p. 298–303) Prof. Pearson has dealt with the same problem, basing the values of the constant upon the successive summations of $\log l_x$.

It is, perhaps, preferable to deal directly with the actual exposures and deaths in a manner similar to that first described by Makeham (*J.I.A.*, vol. xvi, p. 344). This can be readily done, and the same method of summations or moments applied as in the case of any other frequency curve.

Tabulate $E_{x+\frac{1}{2}}$, that is, the number exposed to risk in the middle of the year of age x, and θ_x representing the deaths occurring between ages x and $x+1$. Assuming, as we may with sufficient accuracy for ordinary purposes,* that the force of mortality at age $x+\frac{1}{2}$, or the function colog $_e p_x$, is equal to m_x the "central death rate" $=\dfrac{\theta_x}{E_{x+\frac{1}{2}}}$, we have

$$E_{x+\frac{1}{2}}(A+Bc^{x+\frac{1}{2}})=\theta_x.$$

If we knew the value of c, we could then tabulate the values of $E_{x+\frac{1}{2}}$, $E_{x+\frac{1}{2}}c^{x+\frac{1}{2}}$, θ_x respectively, and summing these values continuously to the end of the table, and again taking the total of these sums, we should obtain equations in this form :—

$$(\Sigma E_{x+\frac{1}{2}})A+(\Sigma(E_{x+\frac{1}{2}}c^{x+\frac{1}{2}}))B=(\Sigma\theta_x)$$

$$(\Sigma\Sigma E_{x+\frac{1}{2}})A+(\Sigma\Sigma E_{x+\frac{1}{2}}c^{x+\frac{1}{2}})B=(\Sigma\Sigma\theta_x)$$

a simple simultaneous equation for determining A and B. As a matter of fact, the value of $\log_{10}c$ does not usually differ very much from ·04, and in general it will be found that a small change in the value of log c does not involve a serious change in the general character of the table. In an important series of observations, however, we cannot assume the value of c. Either we must determine c by a method such as that used by Mr. Woolhouse or Mr. King, which will give a sufficient approximate value, or we may adopt two or more alternative values of c, which appear likely to contain between them the true value. Having obtained the values of constants A and B for each given value of c, set out the expected or graduated deaths, and compare them with the actual numbers in suitable age groups. If the third summation of the differences of the graduated and ungraduated deaths is computed, it will be possible by

* Assuming the usual table of E_x and θ_x to represent accurately the facts and to be undisturbed at the older ages (where alone the point is of any importance) by entrances or by exits other than by death, then $\dfrac{\theta_x}{E_x}=q_x$ acurately; and colog $_e p_x=\dfrac{\theta_x}{E_x-\frac{1}{2}\theta_x-\frac{1}{12}m_x\theta_x}$, very nearly, where m_x is the "central death rate" $\dfrac{\theta_x}{E_x-\frac{1}{2}\theta_x}$. The error caused by omitting the small term in the denominator and taking colog $_e p_x=\dfrac{\theta_x}{E_x-\frac{1}{2}\theta_x}$ is only appreciable at the older ages, amounting to 1 per-cent in the rate of mortality where $q_x=·3$ or about age 90.

interpolation to obtain a value of $\log c$, making these nearly equal to zero. Putting the matter into the language of moments, we shall then have made the first, second and third moments of the graduated and ungraduated curves equal, and in that way we shall have selected what may be considered the best values of the constants A, B and c.*

It may be objected that the use of this particular method is open to the same implication of giving equal weight to all the observations, as in the case of the values of l_x. We can avoid that objection by duly weighting the observations at each age by multiplying the "exposed" and "died" at each age by the approximately graduated values of $(\theta_x)^{-\frac{1}{2}}$. But although this would give suitable weights to the observations, if the curve of mortality were a parabolic curve, or if it were known to follow accurately Makeham's Law, it is not quite clear that it would do so in practice. It may be assumed that (when the constants are formed by reproducing the moments of the deaths) in not weighting the observations, we give less weight to those at the commencement and at the end of the table than they are theoretically entitled to. But this is not a serious practical objection. Makeham's law is only approximately correct, and as we reach younger adult ages it begins to diverge from the facts of observation; on the other hand, as we reach the older ages the actual importance of the observations is less than the weight to which they are theoretically entitled, as estimated by the number of deaths, owing to the fact that the actual mortality at those ages does not materially affect financial questions such as rates of premium and reserves.

Beyond this consideration there is also a degree of doubt attaching to the rates of mortality at extreme ages in any table.† Indeed, we may go further, and say that in all considerable tables of statistics the numbers at the extremes of the table are proportionately more affected by sporadic or accidental errors of observation than those in the body of the table. If we suppose that in a very small percentage of cases the ages of the "Exposed to Risk" and "Died" are affected by errors of calculation, clerical errors in transcribing the data, &c.—these cases being removed from their true position and scattered at random over the table—the

* *See* Note G, p. 131.

† *See* my notes on this subject in "Principles and Methods", p. 148.

effect upon the data over the great bulk of the table will be insignificant owing to the large numbers under observation and to a balance of errors, but the effect upon the experience at the extremes of the table, where the actual numbers under observation are very small, may well be appreciable.

Reverting to the problem of obtaining the value of c in Makeham's formula directly from the observations, we may endeavour to represent the curve of the "Exposed to Risk" by some frequency curve which can be suitably combined with the formula for μ_x to represent the deaths—such, for example, as the normal curve $y=ke^{-x^2/c^2}$, or the curve No. 5, $y=kx^ne^{\gamma x}$, or by the terms of a binomial expansion (*see* Calderon, *J.I.A.*, vol. xxxv, p. 157). Unfortunately none of these curves give a very satisfactory representation of the average form of the "Exposed to Risk" curve. In the case of the binomial, in order to get a tolerable fit, it will be generally found that the value of n in the expression $\dfrac{ka^x}{\lfloor x \, \lfloor n-x}$ (representing the general term of the binomial) must be taken small; that is to say, the data must be arranged in somewhat large groups of not less than about 10 ages to a group. In either case it will be necessary, after obtaining a frequency curve fitting the numbers of the "Exposed to Risk," to re-compute the deaths on the basis of these graduated numbers.

Thus, while it is possible to determine the values of c directly from the observations, the process is laborious. In my opinion, it is preferable to use certain trial values of c which we know to lie near the truth, and, by a comparison of the resulting graduated deaths with the original facts, to select a value which appears to give the best general agreement, which may not always be that making the third summation of the deviations zero.*

There is a further point to be considered with respect to the nature of the differences between the original numbers, whether of deaths or of other observations, and the numbers obtained by a graduation following a formula such as that of Makeham. These divergences between the ungraduated and graduated numbers will in part arise from the smallness of the numbers under observation, and may in part arise from the fact that the formula does not accurately represent

* *See* Note G, p. 133.

the true curve of mortality. For the majority of mortality tables, for male lives at the adult ages, Makeham's formula is so near the truth that we may in practice neglect the systematic errors and assume that the formula represents the true curve of mortality, determining our constants as though the whole of the deviations in the graduated and ungraduated curves are accidental and due to the smallness of the data, but for some tables, notably those representing the mortality of females, this will not be the case.

Other expressions may be given representing approximately the curve of μ_x, as, for instance,

$$\mu_x = ma^x + nb^x \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

whence

$$\log_{10} l_x = \mathrm{K} + \mathrm{M}a^x + \mathrm{N}b^x \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

an expression which enables us to represent some mortality tables, such as those arising from tropical experience, that are not very readily represented by Makeham's formula. The values of these constants can be readily obtained either from 5 selected values of log l_x or from the sums of the values of selected groups of the same function.

The above formula for l_x preserves in a modified form the principle of uniform seniority. Not, however, in a very practicable shape as in order to compute values of joint-lives (any number) we require tables of k joint-lives of equal age for various values of k. It is of course evident from general considerations that the force of mortality on any number of joint-lives must consist of two terms, each of which is a member of a geometrical progression, and that if we can find an age w where the *relative* values of these two terms is the same as in the joint-life status, the actual values will be the same when multiplied by some suitable constant k. The required joint-life annuity will then be represented by the annuity on k joint-lives all of age w.

Take as an example an annuity on the joint-lives of (x) and (y). Find k and w so that

$$\left.\begin{aligned} a^x + a^y &= ka^w \\ b^x + b^y &= kb^w \end{aligned}\right\} \begin{aligned} &\text{whence } w = \frac{\log(a^x + a^y) - \log(b^x + b^y)}{\log a - \log b} \\ &\text{and} \qquad k = (a^x + a^y) \div a^w = (b^x + b^y) \div b^w \end{aligned}$$

Then it is obvious that if we replace x and y by $x+t$ and $y+t$, k will remain unaltered and w will become $w+t$, so that the

principle of uniform seniority is maintained. Thus, an annuity on x and y will be equal to an annuity on k lives all aged w; or, since k will not generally be integral, it will be more convenient to say that $a_{xy}=a'_w$ where a' is calculated at forces of mortality always k times the normal force, age for age. Thus, we shall require tables for various standard values of k, and we shall usually require a double interpolation, since neither w nor k will usually be integral.

The principle of employing the sum of two (or more) geometrical series to represent the logarithm of a function such as the number living may also be used with advantage, as will be seen later on, for census tables. (*See* the Sixth Lecture.)

As an example of this formula, we may apply it to the column of $\log l_x$ in the O^{M} Table.

Taking the values of $\log l_x$ for ages 20, 37, 54, 71 and 88, we have the following data:

$$\log l_{20}=4\cdot98432=K+Ma^{20}+Nb^{20}$$

$$\log l_{37}=4\cdot94279=K+Ma^{37}+Nb^{37}$$

$$\log l_{54}=4\cdot85300=K+Ma^{54}+Nb^{54}$$

$$\log l_{71}=4\cdot58086=K+Ma^{71}+Nb^{71}$$

$$\log l_{88}=3\cdot47509=K+Ma^{88}+Nb^{88}$$

whence differencing, and writing

$$Ma^{20}(a^{17}-1)=-M';\ Nb^{20}(b^{17}-1)=-N';\ a^{17}=\alpha;\ b^{17}=\beta;$$

we have

$$M'+N'=\log l_{20}-\log l_{37}=\ \cdot04153=A$$

$$M'\alpha+N'\beta=\log l_{37}-\log l_{54}=\ \cdot08979=B$$

$$M'\alpha^2+N'\beta^2=\log l_{54}-\log l_{71}=\ \cdot27214=C$$

$$M'\alpha^3+N'\beta^3=\log l_{71}-\log l_{88}=1\cdot10577=D$$

whence, noting that

$$\frac{BD-C^2}{AC-B^2}=\alpha\beta;\ \frac{AD-BC}{AC-B^2}=\alpha+\beta;$$

we easily obtain:

$$\alpha = 5{\cdot}1082 \quad ; \quad a = 1{\cdot}1007$$

$$\beta = 1{\cdot}5243 \quad ; \quad b = 1{\cdot}0251$$

$$M' = {\cdot}0073886 ; \quad M = -{\cdot}00026403$$

$$N' = {\cdot}0341414 ; \quad N = -{\cdot}039657$$

The following comparison of the values of l_x and decrements for quinquennial ages will indicate the approximation of the formula to the O^M Table.

TABLE XI.

Values of l_x and of $(l_x - l_{x+5})$ according to the O^M Table, as compared with re-graduation by formula (2).

Age	l_x		Quinquennial Decrements			
	By Formula	Original Value	By Formula	Original Value	Errors	
					+	−
20	96,453	96,453	2,129	2,066	63	...
25	94,324	94,387	2,467	2,445	22	...
30	91,857	91,942	2,896	2,947	...	51
35	88,961	88,995	3,443	3,528	...	85
40	85,518	85,467	4,158	4,205	...	47
45	81,360	81,262	5,108	5,077	31	...
50	76,252	76,185	6,350	6,266	84	...
55	69,902	69,919	7,927	7,846	81	...
60	61,975	62,073	9,775	9,766	9	...
65	52,200	52,307	11,606	11,692	...	86
70	40,594	40,615	12,753	12,863	...	110
75	27,841	27,752	12,192	12,222	...	30
80	15,649	15,530	9,244	9,171	73	...
85	6,405	6,359	4,827	4,763	64	...
90	1,578	1,596	1,406	1,410	...	4
95	172	186	167	179	...	12
100	5	7	5	7	...	2

FIFTH LECTURE.

ALTHOUGH in the preceding Lecture the application of Makeham's formula has been considered at some length, its importance is such that we may now touch on some further points, and particularly on the application of the formula to the graduation of select tables.

The suitability of Makeham's formula to the graduation of mortality tables must be judged as we should judge the applicability of any other frequency curve to a given series of observations. That is to say, we must consider whether the observed differences between the graduated and ungraduated values (the computed and actual deaths) fall within what may be properly considered to be the limits of error. For practical purposes, owing to the great convenience attaching to the use of the formula, it is worth while to stretch a point in its favour. Instead, therefore, of merely considering the closeness of the agreement between the actual and computed deaths, we may consider how nearly the ungraduated and graduated monetary functions, such as the values of premiums or annuities, are in agreement. If this agreement is sufficient for our purpose, we are justified in adopting the graduation as given by the formula, notwithstanding the fact that at certain groups of ages the divergences between the graduated and ungraduated deaths may be greater than would be expected from the theory of probabilities. In this connection it is to be noted that our observations relate to past time, and that the quantities we are measuring are all liable to change with time. Hence in a graduation intended to form the basis of tables of annuities or premiums it is sufficient if the general character of the experience is retained without insisting too strongly upon a strict adherence to minor features. This is illustrated by the following table from "Principles and Methods" (p. 162), in

which we may anticipate for the moment the question of the application of Makeham's formula to select tables :

$O^{[M]}$ Whole-Life Participating—Males.

3 per-cent Premiums for £100 Assured.

Age	$P_{[x]}$ Ungraduated	$P_{[x]}$ Graduated	G−U +	G−U −	Sprague's $H^{[M]}$ Select	$H^{[M]}-O^{[M]}-(3)$ +
(1)	(2)	(3)	(4)	(5)	(6)	(7)
20	1·379	1·365	...	·014	1·563	·198
25	1·535	1·551	·016	...	1·703	·152
30	1·779	1·785	·006	...	1·925	·140
35	2·086	2·081	...	·005	2·218	·137
40	2·453	2·457	·004	...	2·602	·145
45	2·952	2·940	...	·012	3·106	·166
50	3·571	3·564	...	·007	3·755	·191
55	4·338	4·377	·039	...	4·635	·258
60	5·413	5·446	·033	...	5·827	·381
65	6·872	6·854	...	·018	7·433	·579
Average	3·238	3·222	·004	...	3·477	·235

Here columns (4) and (5) show how far the graduated select annual premium $P_{[x]}$, for each age at entry, differs from the ungraduated value for the same age, while column (7) shows how far the annual premiums deduced by Dr. Sprague from the H^{M} data (*Journal of the Institute of Actuaries*, vol. xxii, p. 391) differ from the premiums deduced from the $O^{[M]}$ Experience. The average difference between the graduated and ungraduated premiums (irrespective of sign) amounts to ·015 per £100 assured, a quite insignificant amount; whereas the difference between the premiums representing the earlier experience and those of the $O^{[M]}$ Table, representing the experience of 30 years later, are all positive and average ·235 per £100 assured.

Only a part of the differences shown in columns (4) and (5) are due to any systematic difference between the mortality as shown in the $O^{[M]}$ data and that assumed by the formula. Assuming, however, that the entire differences were due to this cause, it will be seen that the changes introduced into the values of the monetary functions by using Makeham's formula are a very small percentage of the actual change that has occurred in the value of these functions during the course of 30 years.

Although, therefore, the differences between the graduated and ungraduated deaths do at certain points somewhat exceed the limits of the errors of observation, we are justified in using the graduated table as a standard for the future.

Each case must, of course, be decided upon its own merits, and while the H^M Experience and the O^M Experience have, with other tables, proved to be amenable to Makeham's formula, the latter cannot be treated as a "law of mortality", to which all tables may be expected to conform. As already stated, its suitability must be tested, as that of any other frequency curve, but with rather more latitude owing to its practical advantages. In particular the formula is not generally suitable for tables representing the mortality of Female Lives.

In the last lecture we considered various methods of determining the constants of Makeham's formula for μ_x best representing a given mortality experience, in particular that depending upon the agreement between the totals of the graduated and ungraduated deaths and of their successive summations. We have so far, however, considered the force of mortality as a function of the age only, so that our results are applicable only to "mixed" tables of mortality, not to "select" tables in which the mortality is treated as a function both of the age of the life and of the duration of the assurance.

The formula owes its value, beyond the incidental advantage that it gives us a very simple and effective method of graduation, to the relation it establishes between the value of an annuity upon joint lives of any age and that of an annuity upon the same number of joint lives of equal age. From the formula for the force of mortality according to Makeham's hypothesis

$$\mu_x = \text{A} + \text{B}c^x$$

it follows that the force of mortality for any number of joint lives, aged, for example, at entry x, y, z, is given by the formula

$$\begin{aligned}\mu_{x+t} + \mu_{y+t} + \mu_{z+t} &= 3\text{A} + \text{B}c^t(c^x + c^y + c^z) \\ &= 3\mu_{w+t}\end{aligned}$$

where $$c^w = \frac{1}{3}(c^x + c^y + c^z)$$

where t represents the period elapsed since the date of entry. As a value of w satisfying this equation can always be found, and is independent of t, it follows that

$$a_{xyz} = a_{www}$$

It is seen that the relation subsisting between the value of w and the values of x, y, z, involves the constant c only, and not the constants A and B; hence, any variation introduced into the values of the constants A and B, having reference to the time elapsed since selection and depending only on t, will not affect the relation between the age w and the ages x, y, and z. We can, therefore, write the force of mortality at age $x+t$ for a life select at age x as follows:

$$\mu_{[x]+t} = \mathrm{A} + f(t) + [\mathrm{B} + \phi(t)]\, c^{x+t} \quad . \quad . \quad . \quad (1)$$

and still retain the relation

$$\mu_{[x]+t} + \mu_{[y]+t} + \mu_{[x]+t} = 3\mu_{[w]+t}$$

when
$$c^{w} = \frac{1}{3}(c^{x} + c^{y} + c^{z}).$$

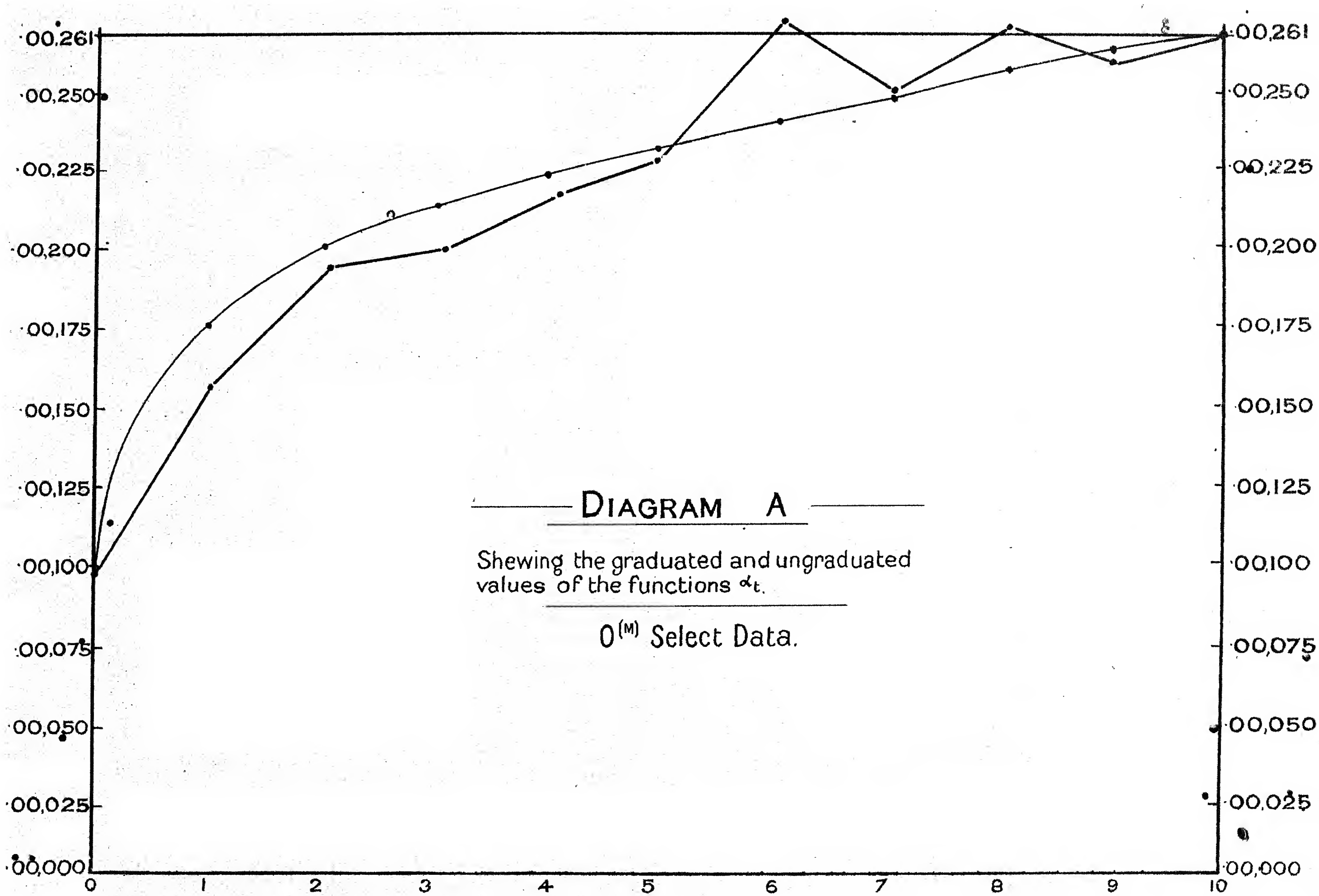
Equation (1) may obviously be written in the form

$$\mu_{[x]+t} = \mathrm{A}_t + \mathrm{B}_t c^{x+t} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

or alternatively, if, as is often more convenient, we work with the values of colog p_x, in the form

$$\operatorname{colog}_{10} p_{[x]+t} = \alpha_t + \beta_t c^{x+t} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where A_t and B_t, or α_t and β_t, may be any functions of t, but are not functions of x. We can thus represent the rate of mortality as a function both of the age and of the time elapsed since selection and so approximate fairly to the rates of mortality shown in an "analyzed" or "select" mortality experience, while retaining most of the advantages arising from the use of Makeham's formula. The two functions of t have probably a tendency to become constant as t increases but do not necessarily become so within any special period from the date of entry; they may continue to change slowly throughout the whole duration of the table, and in theory, no


·00,261
·00,250
·00,225
·00,200
·00,175
·00,150
·00,125
·00,100
·00,075
·00,050
·00,025
·00,000
0
1
2
3
4
5
6
7
8
9
10
DIAGRAM A
Shewing the graduated and ungraduated values of the functions α_t.
O[M] Select Data.

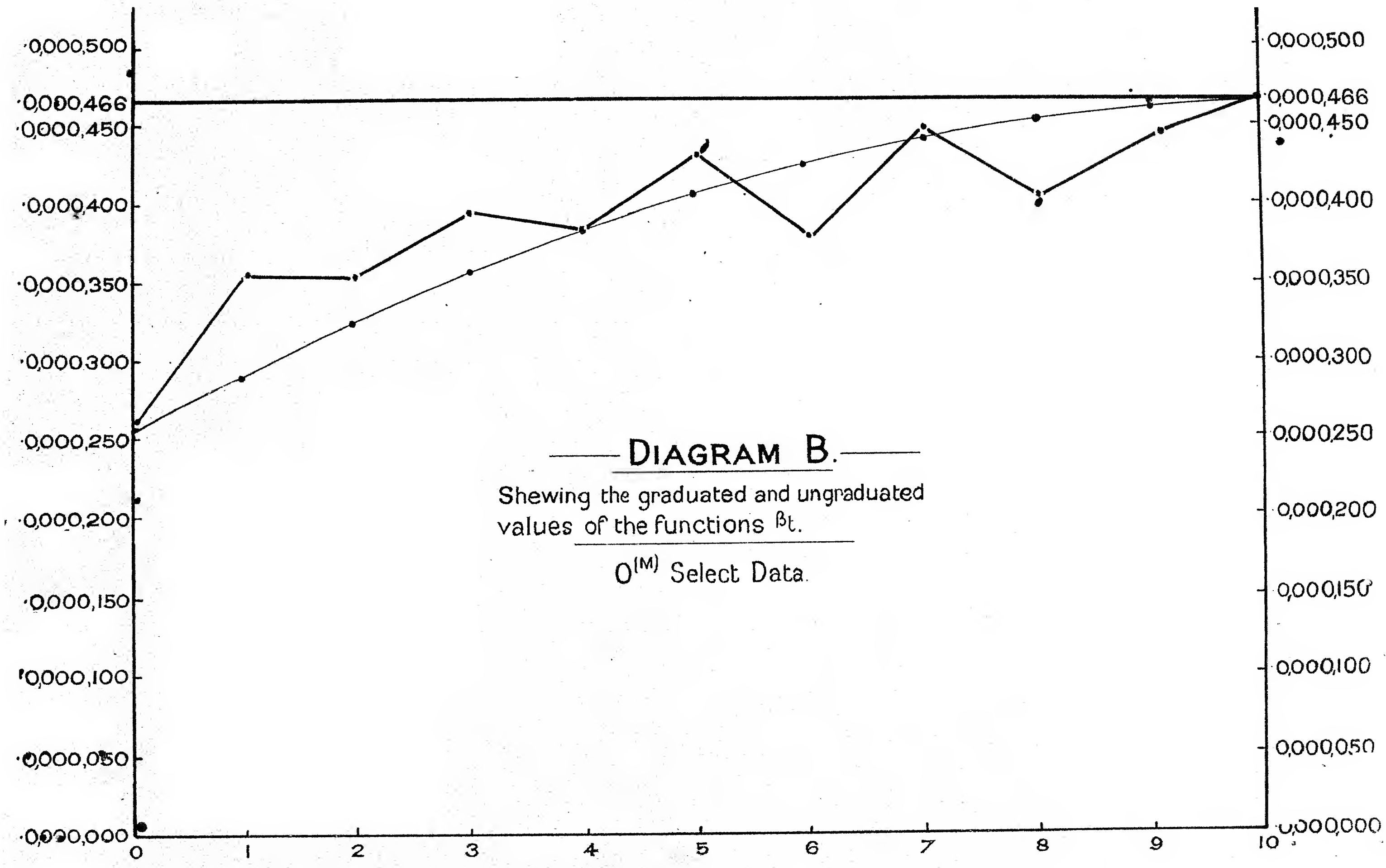


DIAGRAM B.

Shewing the graduated and ungraduated values of the functions β_t.

$O^{(M)}$ Select Data.

doubt, should do so, but for practical purposes it is convenient to make them constant after a few years (say 5, or at most 10) from the date of entry, beyond which point it is assumed that the effect of "selection" has worn off.

If we set out separately the data for each year of assurance, that is, for each value of t so far as we intend to trace selection, we shall have a series of equations (corresponding to those shown on p. 65 for an aggregate table) for determining the numerical values of the functions $f(0)$, $f(1)$, &c., $\phi(0)$, $\phi(1)$, &c., the value of c being necessarily that determined for the "ultimate" table. In other words, the data for each year of duration are treated as representing a mortality table complete in itself. We obtain in this way values for A_t and B_t or for α_t and β_t for each value of t, so far as it is proposed to carry the select tables. Unless, however, the experience is a very large one, these values will be very irregular. Indeed, in the case of the O^M data, which represent a large experience, we have somewhat irregular values for α_t and β_t, even during the first ten years of assurance, where the facts are most numerous. The approximate values of α_t and β_t for the O^M data are given on p. 157 on "Principles and Methods." If these values are plotted out, the resulting curves exhibit certain obvious characteristics, as will be seen by the diagrams opposite where the regular lines show the ungraduated, and curved lines graduated values of α_t and β_t, and the horizontal lines after 10 years represent the values for the experience of 10 years' duration and upwards, when they are assumed to be constant. A period of 10 years would appear from the data to be the shortest within which we can effect anything like a smooth junction between the "select" and "ultimate" mortality rates.

The values of α_t rise very rapidly in the first few years of assurance, but after about 6 or 7 years they appear to approach nearly their final value. In the case of β_t, however, we see that if the graduated curve were drawn as closely as is consistent with smoothness through the ungraduated values, it would probably not reach the level of the ultimate value ·0000466 until after 15 years from entry, and even then it would be below the value of β_t for durations of 15 years and over. Hence it would seem that the value of β_t does not become constant until about 20 years have elapsed from the date of entry. We may almost say that while the effect of

selection as reflected in the values of the α constant disappears after about 7 years, the effect upon the values of β probably continues throughout the whole of life. The explanation is, no doubt, that the α or A constant represents mortality from accidental causes and from non-constitutional diseases of short duration, whereas the β or B constant represents mortality due to diseases of longer duration and to constitutional defects.

Having obtained numerical values of α_t and β_t for successive values of t, it remains to represent these values by convenient formulæ. The fact that the function β_t does not reach its ultimate value at the end of 10 years from entry, involves either some sacrifice of the agreement between the adjusted and unadjusted values of this function, or a continuation of the analyzed mortality rates beyond the period of 10 years, which is not very convenient. In consequence of this fact we cannot apply the method of moments in fitting a graduated curve to these values. Where the fitting of a frequency curve involves any systematic departure from the original facts, the method of moments often gives unsatisfactory results, and a curve may be produced departing more widely from the observations than if derived by a tentative method.

In selecting formulæ for graduating the rough values of α_t and β_t, there are certain conditions which should be fulfilled:

1. A smooth junction between the curves representing the select and ultimate tables.

2. An agreement between the graduated and ungraduated values of α_t, β_t in year 0, as a special importance attaches to the rate of mortality in the first year of assurance.

3. An agreement between the aggregate graduated and ungraduated values of these functions during the period between the date of entry and the ultimate table.

To conform to these conditions as far as possible, we must select a curve for the values of β_t which, whilst running smoothly into the constant value at the end of ten years, will represent fairly well the distinctly lower values of β_t in the years immediately preceding. This may be done by representing the difference between $\log l_{x+t}$ (the value of this function in the "ultimate" table) and $\log l_{[x]+t}$ (the value in the "select" table) so far as this difference is due to changes

in β_t, by an expression of the form $n(10-t)^2\beta c^x$, where β is the ultimate value; whence we have the corresponding difference:

$$\mu_{x+t}-\mu_{[x]+t}$$

$$=2n(10-t)\beta c^x$$

$$=2n(10-t)c^{-t}\beta c^{x+t}$$

so that $$\beta_t=[1-2n(10-t)c^{-t}]\beta.$$

The result of this is to eliminate from the β constant at the latter durations part of the effect of selection, and somewhat to exaggerate the effect in earlier years.

We have now to decide as to the curve best representing the values of a_t. The method employed will depend very much on the character of the experience we are treating. In the $O^{[M]}$ Experience it was again found convenient to adopt an expression for the difference of $\log_{10}l_{x+t}$ and $\log_{10}l_{[x]+t}$, so far as this difference was due to change in a_t, containing a term similar to that due to β_t with the addition of a further term representing a geometrical series rapidly diminishing as t increased. The final form of the equation for the $O^{[M]}$ Experience was as under—

$$\log_{10}l_{[x]+t}=\log_{10}l_{x+t}-m(10-t)^2-m'(c')^t-n(10-t)^2\beta c^x.$$

Having determined the form of this equation, the simplest method for determining the constants is to express in terms of them the difference between the computed deaths by the ultimate table of mortality, and the actual deaths for each age or each group of ages and each year of assurance.

We have in that way a series of equations for determining the values of these constants m, m', c', n, and hence of A_t and B_t for each value of t, similar in principle to the equations used for determining the values of the original constants A and B. The only point that arises is as to what particular way we are to group the observations to determine those values.

The value of m in the above formula having been ascertained with a view to representing as nearly as practicable the effect of selection upon the constant B, there remain in all, four unknown quantities in the formula to be determined, and the actual equations used to

determine them were formed by taking the first and second summations by ages of the whole of these expressions, representing the difference between the "select" and "ultimate" rates, first for year of assurance 0 alone, and then for the whole of the ten years.

The selection of these particular groups is, of course, not a question of principle, but of convenience. Each case must be treated with reference to the nature of the curve of selection, as brought out by the statistics, and such a process adopted as appears to be calculated to bring out the best results in the particular case in question.

It may happen in certain tables that it is inconvenient to trace out the effect of selection for so many years, and in particular this is the case in a table representing the mortality of annuitants. In such a table the effect of selection (which is here the self-selection of the annuitant) persists for a very long time. In a table of insured lives, owing to the cessation of new entrants in middle life, practically at about age 55, the mortality at the older ages is but slightly affected by selection. In the case of annuitants, where there is a constant inflow of fresh lives up to 75 or 80 years of age, the mortality is affected by this cause throughout the whole extent of the table. To completely represent the effect of selection in such an experience will require an elaborate series of tables, showing for each entry age the value of annuities for each year elapsed since entry for many years duration. The tables given in "Principles and Methods", pp. 124, 125, show that as regards the O^{am} and O^{af} Experience, and doubtless the same feature would be found to be general, the values of the expectation of life ten years after entry are appreciably greater than the values for the same ages derived from the "ultimate" rate of mortality ($e_{[x]+10} > e_{x+10}$). Consequently, if the graduated rates of mortality for the first five or ten years from entry are employed in conjunction with rates representing the aggregate mortality after five or ten years, as the case may be, the ultimate values of the annuities, and also the values of the date of entry will on the whole be under-estimated. In any table used for the grant of annuities it is, however, most important that annuities at the date of entry shall not be undervalued, and of only less importance that the values in succeeding years shall be such as may be safely employed in estimating reserves. Any method,

therefore, of treating an annuity experience which tends to underestimate the values of annuities is clearly unsuitable. Full weight must accordingly be given to the effect of selection, but to avoid the heavy work involved in a complete analysis, the expedient may be adopted of computing a hypothetical table of mortality which will correspond to the values of the annuities, let us say, five years from the date of entry. If this can be done successfully and the rates of mortality for the first five years joined on smoothly with the rates in such hypothetical table, we shall then have a correct measure of the value of annuities at entry and for the five years following, while thereafter the values will be slightly, but not seriously, overestimated, an error which will be on the right side.

We may take as our basis either the values of the "expectation of life" or of the annuities at a suitable rate of interest. We will assume the former to be adopted. As these values ($e_{[x]+5}$) will depend upon separate groups of data, viz., the entrants at individual ages, it will not be practicable to construct an ungraduated table of p_x from the formula $p_x=\frac{e_x}{1+e_{x+1}}$ the irregularities in the individual values of e_x leading to anomalous results. A better plan will be to graduate the table of expectations. For this purpose, we may assume any frequency curve which will represent these expectations satisfactorily, for example, a curve such as $\log_{10}e_x=a+bx+cx^2+dx^3+fx^4$. We may employ values of e_x deduced from the experience of individual ages at entry, or we may combine the entrants in quinary groups of ages, taking due account of the true average age of each group of entrants.

The only point of importance where difficulty arises is the weighting of the different equations. These are not of equal weight because the expectations of life, as deduced from the unadjusted experience, are based upon a smaller or larger experience, as they fall at the extremes or in the middle of the table, and some method must be devised for giving due weight to this fact. This may be done by simply weighting the equations with the actual number of entrants at that particular age, and much may be said for this method although it slightly underestimates the weights at the extreme ages. If we are dealing, for example, with the

values of annuities, and approximately the same result will be arrived at when working with the expectations of life, the plan of weighting the unadjusted values in proportion to the number of lives entering at each age, would make the total cost of all the annuities by the graduated table the same as by the ungraduated, an agreement that would have some practical value. In the alternative, we may consider that each value of the expectation of life (or of the annuity, as the case may be) should be weighted in proportion to the reciprocal of its average error. Thus if $e_{[x]+5} \simeq A \pm z$, where A is the observed value and z the average error, we shall have $\frac{e_{[x]+5}}{z} = \frac{A}{z} \pm 1$. It is difficult to determine satisfactorily the average error in the value of the unadjusted expectation of life,* the problem being complicated by the incompleteness of the observations due to the "existing." A fairly satisfactory method of estimating the average error would be as follows. Taking the series consisting of the values of e_x for all values of x, each of those values depending on a given age at entry only, we may assume that the observed second differences of these quantities $e_{x-1} - 2e_x + e_{x+1}$, which, in a well graduated table, would be very small, are due to the errors of observation in the values e_{x-1}, e_x, and e_{x+1}. In any particular group of entry ages, we may say that the average of the central second differences (taken irrespective of sign) will be, on the average, proportional to the average error in e_x for that particular group.† Computing the average values of the central second differences (without sign), for various sections of the table, and drawing a smooth curve through them, we should obtain values from which suitable relative weights for the individual observations could be deduced.

This would be a very fair method of determining practically the weight to be attached to the values of e_x in different parts of the table. Or we may proceed, as was actually done in the case of the annuity experience graduation, by assuming the error in the value of e_x to be a function, first, of the total number of deaths in the experience representing the particular entry age, and secondly, of the age x. This method may appear somewhat arbitrary, but as only the *relative* weights

* *See*, however, the Sixth Lecture, pp. 100–104.

† The average value of $e_{x-1} - 2e_x + e_{x+1}$ will be $\sqrt{6}$ times the average error in e_x.

are in question, it is sufficient for the purpose. It must be understood that the relative weights adopted do not very greatly affect the results. The values of Makeham's constants as deduced, for example, from the values of $\log l_x$ for ages 25, 45, 65, 85, thus giving equal weight to the observed value of mortality from ages 25 to 85, would not generally differ materially from the values resulting from a careful system of weighting, although, of course, the latter are to be preferred.

Assuming the "exposed to risk" to remain unchanged, the average error in the observed number of deaths is approximately $\pm \cdot 8\sqrt{nq(1-q)}$ where n is the total of the "exposed to risk" and nq the total deaths. The average percentage error in the total deaths will, therefore, be proportionate to $\pm\sqrt{\frac{1-q}{nq}}$. If we suppose that this average error is distributed uniformly through all ages passed through by the particular group of entrants, we can then arrive at a rough estimate of the average error in the observed value of e_x, by computing the effect of a change of, say, 1 per-cent in the mortality rates throughout.

The assumptions here are not strictly accurate, as errors in the value of e_x arise not only from the total number of deaths being greater or less than the expected amount, but from the manner in which the excess or defect of mortality is distributed through the table. The neglect of this second source of error will not, however, seriously affect the relative weights arrived at, and for practical purposes the relative average errors in the value of e_x will be dependent, first, on the average error in the total deaths observed in the experience from which it is deduced, and second, on the extent to which a given percentage error in the mortality distributed uniformly through the table will affect the value of e_x. The product of these two factors may be taken as representing sufficiently approximately the expected error in the value of e_x, remembering always that this estimated error is not an absolute, but a relative measure at the various ages. When this is done, we have, by taking the reciprocals of those quantities, the weights which we shall give to the observed values of e_x in order to determine our constants.

It is necessary to point out that this process, while suitable for expectations calculated from entrants at a particular age

or small groups of ages, will not apply to aggregate tables; for in their case the percentage error in the total deaths above age x steadily increases as x increases, so that this method would produce weights steadily diminishing from the youngest age to the oldest, which would obviously be incorrect.

Notwithstanding the important effect of selection on mortality, it is frequently ignored, as in the H^M and O^M Tables. It is important to consider, therefore, what is the net effect in a mortality table of neglecting altogether the factor of selection. Considerable additional labour attaches to the use of select tables for valuation purposes, and the question may be asked what kind of errors do we make if we neglect the fact that mortality is a function not only of the age, but also of the duration of assurance, and treat it simply as a function of the age as it is treated in the O^M and H^M Tables. In a mortality table representing assured lives the effect will be seen if we compare a table like the H^M Table with a table like Dr. Sprague's Select Table, or if we compare a table such as the O^M Table with a table like the $O^{[M]}$ Select Table:

Comparison of Annual Premiums for the Assurance of 100 (3 *per-cent interest.*)

Age	H^M	$H^{[M]}$ Sprague	O^M	$O^{[M]}$
20	1·427	1·563	1·306	1·365
25	1·625	1·703	1·524	1·551
30	1·880	1·925	1·790	1·785
35	2·193	2·218	2·116	2·081
40	2·589	2·602	2·524	2·457
45	3·114	3·106	3·046	2·940
50	3·801	3·755	3·730	3·564
55	4·725	4·635	4·641	4·377
60	5·987	5·827	5·872	5·444
65	7·705	7·433	7·557	6·853

If we compare, as is most convenient, either annuity or premium-values, we shall find that the effect of ignoring the element of selection and treating the mortality rates as a function of the age alone is that, at the younger entry ages, premiums are underestimated and annuity-values are

overestimated.* The $O^{[M]}$ premiums should, properly speaking, be compared with those derived from a table representing the true aggregate of the select tables, but no such table is available. There is a point, which is in general somewhat greater than the average age at entry, at which the two curves representing the premium values for the mixed and select data cross each other, and for the older ages the premiums by mixed tables are greater than those by the select table. The extent of the differences in the premiums is sufficient to render it necessary, in adopting a basis for assurance premiums, to take into account the question of selection. The only plan by which the use of select tables can safely be avoided, is either by adopting a special form of loading or by throwing out altogether from the data upon which the premiums are based those years of assurance which are seriously affected by selection, that is to say by employing a table of the $H^{M(5)}$ or $O^{M(5)}$ type. We then obtain a table which at all ages overestimates the values of the premiums and underestimates the values of annuities.

A table representing "ultimate" rates of mortality, that is, of the $H^{M(5)}$ or $O^{M(5)}$ type, is therefore a safe one to employ for the grant of assurances, although not for the grant of annuities. There is, indeed, very much to be said for the use of a table of that kind for assurance purposes, but, to discuss that question, we should have to go into the finance of life assurance valuations, which hardly comes within the scope of our subject.

With a view of avoiding the necessity for select tables, a device was adopted by the American offices in their first experience denominated the "final series" method. The object was to produce a table not entirely unaffected by selection, but in which its influence would be reduced to a minimum; a table of mortality similar to that which might be supposed to prevail in an office of great age doing a uniform and steady new business. To produce that result the lives

* This is shown, in the table above, to be the case both with the H^M and O^M Tables. Unfortunately, however, in neither case is the comparison very satisfactory. Dr. Sprague's $H^{[M]}$ premiums from the method of their calculation are probably somewhat higher than the true values, and in the case of the O^M Table we are comparing select premiums based in part upon the aggregate of the select tables, excluding first ten years from entry, with O^M premiums based upon an aggregate table from which there had been a further elimination of duplicate assurances.

existing at the close of the observations were traced out through a hypothetical future in which they were assumed to be subject to rates of mortality and lapse identical with the rates actually observed in the past among lives of similar age and duration. The minor details of the process we may pass over. The result from a financial point of view is that the premiums are still underestimated for the younger insuring ages, although not to the same extent as in a table of the H^M type, and are overestimated at the older ages, the point at which the values cross the true curve being earlier than would have been the case had the "final series" adjustment not been used. There are some practical difficulties in adopting a method of this kind. One of these is that after some 15 or 20 years' duration the observed rates of mortality for individual ages and years of assurance depend on a very few facts. We then have to apply the very irregular rates resulting from those few facts to much larger numbers, including the existing lives that have been brought back hypothetically under observation; so that where these irregularities become inconveniently large, the application of the method must cease; or else these irregular rates must be subjected to some process of graduation before being used in the calculations.

This difficulty could be met by using a species of $O^{M(15)}$ or $O^{M(20)}$ Table for risks of 15 or 20 years' duration and upwards, instead of the rates of mortality deduced from individual years of assurance. There are, however, other objections to this method as an expedient for counteracting the effect of a too short average duration of assurance.

As the rate of mortality amongst assured lives cannot strictly be treated as a function of age alone, but is also dependent upon the duration of assurance, so the rates of sickness in a Friendly Society, or of re-marriage in a Widow's Fund, are affected, respectively, by the duration of membership, or of widowhood. Sufficiently approximate results may, however, be generally arrived at in these cases by treating the rate of sickness, or of re-marriage, as a function of the age alone: in the former case because the effect of selection is not very great and is soon exhausted, in the latter case because the average constitution, as regards the duration of widowhood, of a group of lives passing under observation at a given age will be found to remain fairly

constant (unless the Pension Fund is of recent establishment) and the financial effect of a marriage when it occurs is a function of the age only.

Where, however, we are dealing with rates of discontinuance or lapse, it is important that these should be analyzed both as respects age and duration. Owing to the fact that the financial effect of a discontinuance is mainly dependent upon the duration of assurance, very erroneous conclusions may be deduced by treating the rates as functions of the age alone as has sometimes been done. If this course is adopted special precautions must be taken, such, for example, as deducing the rates from a body of lives representing the "existing" some 10 or 20 years back, and excluding from the "exposed to risk" all more recent entrants, as proposed by Mr. A. W. Watson (*J.I.A.*, xxxv, 313–4).

SIXTH LECTURE.

In the concluding Lecture we shall deal with some miscellaneous points of general interest or arising out of the previous Lectures. We have already dealt with the nature of the modifications of Makeham's formula for the force of mortality, necessary to enable us to represent satisfactorily the mortality shown by select tables such as the $O^{[M]}$. These modifications consisted in treating the quantities A and B or a and β in the formulas

$$\mu_{[x]+t}=A+B.c^{x+t};\ \operatorname{colog}_{10}p_{[x]+t}=a+\beta.c^{x+t}$$

which are constants as regards the variable x, as functions of t the time elapsed since the date of selection.

It is clear that a similar course may be pursued if any other formula than Makeham's is employed in the graduation of the "ultimate" table. Thus we may write

$$\mu_{[x]+t}=A_t+B_t.\mu_{x+t}$$

where A_t and B_t will in general be such functions that, as t reaches a certain value, at which the select and ultimate mortality rates merge, A_t becomes zero and B_t unity. The form of these expressions employed for representing the effect of selection suggests that a similar form may be employed for representing rate of discontinuance, which in general may be taken to be a function of the duration of assurance and of the age at entry. The same remark applies to such a function as the rate of remarriage amongst widows, which is, similarly, a function of the duration of widowhood and of the age.

Although we have dealt at considerable length with the use of Makeham's formula in connection with mortality tables, there are some further remarks to be made as to its employment in certain special cases, more particularly in connection with the age statistics at a Census.

If we suppose a population which is (1) subject to uniform rates of mortality, corresponding at the adult ages to Makeham's formula, (2) such that the numbers living represent the survivors from a number of births increasing annually in a geometrical progression, and (3) is subject to a rate of emigration or immigration uniform at all ages, then if l'_x represent the numbers in the population, at a given moment of time, passing through the exact age x, obviously the curve of l'_x will follow Makeham's formula, and if we write

$$\mu'_x = \frac{\frac{d}{dx} \cdot l'_x}{l'_x} = (\mathrm{A}+r) + \mathrm{B} \, . \, c^x$$

we shall have a formula similar to the usual formula for the force of mortality, but with the constant A increased by r, the rate per annum at which the population is increasing; that is to say, the "natural" rate of increase less the rate of emigration. It is true that hardly any population will be found to conform very closely to the above assumptions, but nevertheless it will be frequently found that the population curve for the adult ages does conform to Makeham's formula for l_x, although in most cases it will be necessary to adopt Makeham's second development of Gompertz, with the additional constant in the expression for μ_x.

If the population is given, as is usual, for decennial age groups (*e.g.*, 15-25, 25-35, 35-45, &c.), the values of the ordinate for the middle age of each group may be obtained with sufficient approximation by deducting from each term u_x of the series representing the numbers in successive age groups one twenty-fourth of the central second difference

$$\left(\frac{\Delta^2 u_{x-1}}{24}\right)$$

From the values of l'_x thus obtained, by writing

$$\log l'_x = \mathrm{K} + s.x + g.c^x,$$

or,

$$\log l'_x = \mathrm{K} + s.x + h.x^2 + g.c^x$$

as the case may be, the constants may be determined as for a mortality table.

Take, for example, the male population of England and Wales, enumerated at the Census of 1901, as under :—

TABLE XII.

Male Population in Age-groups: England and Wales, 1901.

Age Group	Numbers* (u_x)	Central Ordinate $u_x - \frac{\Delta^2 x-1}{24}$	log (3)	Δlog (3)	Δ^2log (3)	Δ^3log (3)	Col. (4) Adjusted
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
15–25	94,693						
25–35	76,425	76,373	4·8829	−·1093			4·88349
35–45	59,394	59,371	4·7736	−·1415	−·0322	−·0138	4·77301
45–55	42,924	42,863	4·6321	−·1875	−·0460	−·0485	4·63269
55–65	27,913	27,838	4·4446	−·2820	−·0945	−·0987	4·44401
65–75	14,691	14,541	4·1626	−·4752	−·1932		4·16319
75–85	5,080	4,868	3·6874				3·68681
85–and over	552						

*To reduce the magnitude of these numbers, the figures used are those corresponding to a total population (M & F) of 1,000,000 as given in the Census Report. This, of course, does not affect their relative value nor the form of the curve.

Here, evidently, Col. (6) cannot be well represented by a Geometrical Progression, but with Col. (7) this is possible without very serious changes in the values. This would give a formula corresponding to Makeham's second modification of Gompertz, viz.,

$$\log l'_x = \mathrm{K} + \mathrm{A}x + \mathrm{A}'.x^2 + \mathrm{B}.c^x$$

for the values of the logs of the numbers living at age x, given in Col. (4). As these numbers are only approximate, and our object is merely to show the applicability of the formula as a base line, we may adopt a very simple method of determining the constants, similar to that used by Mr. Makeham in his paper on the Law of Mortality (*J.I.A.*, xiii, p. 338 *et seq.*). If the terms in Col. (4) are alternately

diminished and increased by a quantity z, the quantities in Col. 7 will become

$$-\cdot 0138+8z$$
$$-\cdot 0485-8z$$
$$-\cdot 0987+8z$$

These terms can obviously be made to form a geometrical progression by suitably determining z, and their common *ratio*, found by dividing the sum of the second and third terms by the sum of the first and second, will be equal to $\frac{1472}{623}=2\cdot 363$.

Dividing the sum of the first two terms by $3\cdot 363$ we get $\frac{\cdot 0623}{3\cdot 363}=-\cdot 01853$ as the adjusted first term, giving $8z=-47\cdot 3$ and $z=-5\cdot 9$. Hence the transformed series for Col. (4) is as shown in Col. (8), where the progression accurately follows Makeham's second development.

It is on the whole more convenient to deal with the numbers living above age x rather than the numbers for the decennial age groups.

If we treat the numbers in Table XII in this manner, representing the numbers living above age x by the expression

$$\log Q_x = K + ma^x + nb^x$$

we shall have the results set out in the following table, where the values of the constants have been determined by ignoring the extreme values of $\log Q_x$ at ages 15 and 85, and equating the sums of the values of the above expression to the values of $(\log Q_{25}+\log Q_{35})$, $(\log Q_{35}+\log Q_{45})$, &c., by which means we obtain for the values of the constants

$$\log a = \cdot 006420 \qquad (ma^{25}) = -1\cdot 0582$$
$$\log b = \cdot 035184 \qquad (nb^{25}) = -\ \cdot 007933$$
$$K = 6\cdot 4222$$

The five figure logarithms of Q_x were employed in the calculation, but, owing to the nature of the process, the fifth figure in the graduated column cannot then be relied upon; the logs have therefore been throughout cut down to four figures in the table, which is quite sufficient for the purpose of illustration.

TABLE XIII.

Male Population living above the undermentioned ages.—England and Wales, 1901.

(Based upon figures in preceding Table.)

Age x	Proportional Numbers Q_x	$\log Q_x$	$\Delta \log Q_x$	$\log Q'_x = K + ma^x + nb^x$	$\Delta \log Q'_x$	$\log Q'_x - \log Q_x$ +	$\log Q'_x - \log Q_x$ −
15	321,672	5·5074		5·5059		...	·0015
			− ·1514		− ·1498		
25	226,979	5·3560		5·3561		·0001	...
			− ·1783		− ·1785		
35	150,554	5·1777		5·1776		...	·0001
			− ·2179		− ·2177		
45	91,160	4·9598		4·9599		·0001	...
			− ·2764		− ·2766		
55	48,236	4·6834		4·6833		...	·0001
			− ·3754		− ·3753		
65	20,323	4·3080		4·3080		...	...
			− ·5573		− ·5574		
75	5,632	3·7507		3·7506		...	·0001
			−1·0088		− ·9218		
85	552	2·7419		2·8288		·0869	...

The practical identity of the curves at all ages except 15 and 85, which values were not used in determining the constants, suggests that very accurate results might be obtained by making use of a curve of the above form for interpolation of intermediate values of Q_x.

It has been proposed to employ Makeham's formula to represent the curve of sickness rates at successive ages, and this has been done with a certain degree of success, but the practical advantages of the formula as applied to sickness rates are not very apparent, as it is usually necessary to know not merely the total sickness rate at each age but its division into sickness of various durations, as the number of weeks per annum during the first six months of illness, from the sixth to the twelfth month, after the twelfth month, &c. As Makeham shows (*J.I.A.* xvi, 414), the ratio

$$\frac{\text{Weeks sickness experienced in the year of age}}{\text{Exposed to risk in middle of year of age}}$$

is not a function similar to μ_x but to q_x, since it has a definite limit, namely, 52, or 1 if the sickness is expressed in years in lieu of weeks. Hence if we represent the above ratio by the symbol s_x, we should write

$$\log (52 - s_x) = A + B \,.\, c^x.$$

Where, by the constitution of a society there is no formal superannuation, the sickness benefit continuing throughout life, it is almost invariably the practice of actuaries in using Sickness Tables for the purpose of computing contributions or valuing benefits to assume that the so-called "sickness" will become chronic after a certain age, 70, 75, or 80. In such cases, as the rates of sickness actually employed will generally be much below the maximum of 52 weeks, we may use $\log(N-s_x)=A+Bc^x$. The value of N must be determined by trial.

Mr. King has given an example in the graduation of the values in the Text-book mortality table at the youngest ages of a further application of Makeham's formula, the term Bc^x in the expression for the force of Mortality representing, of course, equally well an increasing rate of mortality as in adult life or a diminishing rate as in infancy and childhood.

In the common case of an asymmetrical series the terms of which become zero, or very nearly so, at each end, the following method of employing the "normal" frequency curve to represent the series will often be found convenient and effective, particularly if the data are presented in the form of a few groups. Let the successive ordinates of the curve be represented by the equation $y=f(x)$; we shall assume the total area of the curve to be unity and the area of curve between the limits $x=\infty$ and $x=t$ will be $\int_t^\infty ydx$. Let us write

$$Y_t=\int_t^\infty ydx=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{z}e^{-t^2}dt$$

so that

$$Y_0=1=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{+\infty}e^{-t^2}dt$$

where z is a function of t, the form of which is to be determined by the data. For most purposes it will be sufficient to treat z as a parabolic function of t, but it will be seen later that there are certain cases in which a different hypothesis as to the form of the function z is to be preferred.

An example will make plain the method of proceeding. Take the O^M data as summarized on p. viii of the volume of Unadjusted Data (Whole-life, Males). In the last two columns of the table there is given the "proportionate distribution per-cent" of the exposed to risk and died. Taking the figures there given we obtain the following tables.

The values of z are found by entering a table of $\frac{1}{\sqrt{\pi}}\int_z^\infty e^{-t^2}dt$ for $+$ and $-$ arguments with the values in the second columns of Tables (XIV) and (XV). We may employ a table such as that given by Woolhouse (*J.I.A.*, vol. xvii, p. 50) or that given on pages 138, 139, at the end of these lectures. Note, however, that in each of these tables the function tabulated is $\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$, say I_x, for $+$ arguments only, so that the total area of the curve from $-\infty$ to $+\infty$ is 2 instead of 1. Hence, if Y_t is $>\frac{1}{2}$ we must put

$$Y_t=\frac{1}{2}+K=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{z} e^{-t^2}dt=\frac{1}{2}+\frac{1}{2}\frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}dt$$

$$=\frac{1}{2}+\frac{1}{2}I_z,$$

so that z takes the value corresponding to the tabular value $I_z=2Y_t-1$. Similarly, if Y_t is $<\frac{1}{2}$ we put z negative and numerically equal to the argument, giving $I_z=1-2Y_t$.

Table XIV.

O^M *Data. Exposed to Risk.*

Age t	Proportion Exposed to Risk above age t $=\frac{1}{\sqrt{\pi}}\cdot\int_{-\infty}^{z} e^{-t^2}dt$	Values of z	Δz	$\Delta^2 z$	$\Delta^3 z$	$\Delta^4 z$	$\Delta^5 z$
0	1·00000	∞					
10	·99991	2·6500	−·7840				
20	·99584	1·8660	−·9574				
30	·90060	·9086	−·6171	·3403			
40	·65989	·2915	−·4741	·1430	−·1973	·0849	
50	·39810	− ·1826	−·4435	·0306	−·1124	·0491	−·0358
60	·18795	− ·6261	−·4762	−·0327	−·0633	·0095	−·0396
70	·05951	−1·1023	−·5627	−·0865	−·0538	−·0070	−·0165
80	·00927	−1·6650	−·7100	−·1473	−·0608		
90	·00039	−2·3750					

* Theoretically the values of z corresponding to a total frequency of 1 and 0 are respectively $\pm\infty$. As however $z=\pm3$ corresponds to Y=·999989 or ·000011, $z=\pm3·5$ to Y=·99999963 or ·00000037, and $z=\pm4$ to Y=·999999992 or ·000000008, it will be seen that any value of z over 3 will sufficiently represent the complete distribution or the zero value, and in practice it would be quite sufficient to insert at the ends of the table any convenient value of z over 3, and consistent with the general run of the intervening terms.

Table XV.

O^M *Data. Deaths.*

Age t	Proportion of Deaths above age t $= \frac{1}{\sqrt{\pi}} \cdot \int_{-\infty}^{z} e^{-t^2} dt$	Values of z	Mean Error of z in last place of decimals	Δz	$\Delta^2 z$	$\Delta^3 z$	$\Delta^4 z$	$\Delta^5 z$
0	1·00000	∞ *						
10	1·00000	∞ *						
20	·99925	2·2450	±192	−·8511	·3190	−·1900	·0717	−·0312
30	·97565	1·3939	± 48	−·5321	·1290	−·1183	·0405	−·0285
40	·88854	·8618	± 28	−·4031	·0107	−·0778	·0120	−·0159
50	·74174	·4587	± 24	−·3924	−·0671	−·0658	−·0039	
60	·53731	·0663	± 23	−·4595	−·1329	−·0697		
70	·28908	− ·3932	± 24	−·5924	−·2026			
80	·08169	− ·9856	± 32	−·7950				
90	·00590	−1·7806	± 82					

* See note at foot of Table XIV on preceding page. It is to be noted, that in lieu of the integral of the normal frequency function, the function $e^z/(1+e^z)$ may be used, leading to a method of procedure similar to that referred to on p. 51.

The column containing the mean error or standard deviation of z in the table of deaths is computed as follows. If the total of the series (in this case the total deaths) is n, and the total above a given point (in this case the number of deaths above age t) is m, then the mean error in m is equal to $\sqrt{\frac{m.(n-m)}{n}}$. From this can be calculated the mean errors of the values in column (2). The change in the value of z corresponding to a given change in the values of $\frac{1}{\sqrt{\pi}}\int_{-\infty}^{z} e^{-t^2}dt$ in column (2) being known from the table of this function we obtain the values in column (4). These standard deviations are not inserted in the table of Exposed to Risk, as the principle upon which the mean errors in the proportionate distribution of the deaths are computed is not strictly applicable to the table of Exposed to Risk, when the latter represent observations spread over a long and continuous period, although it would be applicable if the numbers dealt with represented the exposures in a single calendar year.

If we examine the columns of the successive differences of z in the two tables, ignoring the infinite values of z

corresponding to a total distribution of unity we shall see that they exhibit a remarkable similarity in the nature of their progression, especially from the columns $\Delta^3 z$ onwards. It will also be apparent that a very small alteration of the original values of z in either table would be needed to make the fifth differences constant; that is, we may assume without serious error that

$$z = a + bt + c.\frac{t.(t-1)}{\underline{|2}} + \&c.$$

In order to obtain the closest agreement with the original facts due regard would have to be taken of the weights corresponding to the mean errors in the value of z as given in the table. But we shall obtain results quite good enough for all purposes by the following simple procedure. It will be observed from the values of the mean errors that the values of z for ages 40 to 70 have approximately the same weight, those for ages 30 and 80 have somewhat less weight and finally those for ages 20 and 90 much less.

If we combine the values of z in sets, thus,

$$z_{20} + 3z_{30} + z_{40}\,; \quad z_{30} + 3z_{40} + z_{50}\,; \quad \&c.,$$

with their corresponding numerical values we shall obtain six equations to determine the six coefficients, $a, b, c, \ldots f$. Into these equations the values z_{20} and z_{90} will enter once, the values z_{30} and z_{80} four times, and the remaining values five times. We need not compute the numerical values for all these equations as it will be evident that if we write them down and difference them we shall arrive at the following:

$$\left.\begin{array}{lrl}
5a+5b+c= & z_{20}+3z_{30}+z_{40} = & 7{\cdot}2885 \\
5b+5c+d= & \Delta\,(z_{20}+3z_{30}+z_{40}) = & -2{\cdot}8505 \\
5c+5d+e= & \Delta^2(z_{20}+3z_{30}+z_{40}) = & {\cdot}7167 \\
5d+5e+f= & \Delta^3(z_{20}+3z_{30}+z_{40}) = & -\ {\cdot}6227 \\
5e+5f & =\Delta^4(z_{20}+3z_{30}+z_{40}) = & {\cdot}2052 \\
5f & =\Delta^5(z_{20}+3z_{30}+z_{40}) = & -\ {\cdot}1326
\end{array}\right\}$$

From these equations the values of f, e, d, &c., can be obtained with great facility.

Having obtained a formula for z in terms of t, we can now obtain any term in the series and can also obtain the value of y, the ordinate representing the number of deaths at age x (*i.e.*, approximately between ages $x-\frac{1}{2}$ and $x+\frac{1}{2}$) since

$$y=\frac{1}{\sqrt{\pi}}\cdot e^{-t^2}\cdot\frac{dz}{dt}$$

and
$$\frac{dz}{dt}=\Delta z_t-\frac{1}{2}\Delta^2 z_t+\frac{1}{3}\Delta^3 z_t-\frac{1}{4}\Delta^4 z_t+\frac{1}{5}\Delta^5 z_t.$$

It will generally be sufficient to compute the values of y for decennial or at most quinquennial intervals and to interpolate the resulting values of q_x or m_x for the intermediate ages.

The values of the quantities a, b, c, &c., satisfying the above equations, are

$$a=\ 2{\cdot}24374 \qquad d=-{\cdot}186796$$
$$b=-\ {\cdot}849365 \qquad e=\ {\cdot}067560$$
$$c=\ {\cdot}316624 \qquad f=-{\cdot}026520$$

It may be of interest to give the adjusted values of z and the distribution of deaths corresponding to these which are as under:

TABLE XVI.

O^M *Data. Deaths.*

Adjusted values of z and adjusted distribution of Deaths.

Age	z	$\frac{1}{\sqrt{\pi}}\int_{-\infty}^{z} e^{-t^2}dt$	Last column more (+) or less (−) than corresponding column in Table (XV).	
			+	−
0	6·13645	1·00000	...	...
10	3·69061	1·00000	...	...
20	2·24374	·99925	...	...
30	1·39438	·97569	·00004	...
40	·86166	·88848	...	·00006
50	·45872	·74174	·00000	...
60	·06640	·53740	·00009	...
70	− ·39352	·28893	...	·00015
80	− ·98473	·08188	·00019	...
90	−1·78289	·00585	...	·00005
100	−2·90220	·00002	...	...

The principal objection to this adjustment, paradoxical as it may sound, is that it too closely follows the original facts, the deviations being very much smaller than the probable errors of the observations. This is, of course, due to the fact that we have included too many constants in our formula. A constant fourth difference in the values of z, however, may lead to anomalous results, and a constant third difference makes the errors of adjustment too great. The best plan in such a case would be to adjust the exposures by using a constant third difference, to recompute the deaths to correspond to the adjusted exposures in the 10 year groups and then employ a constant third difference for the graduation of the death curve. Or, as an alternative, an expression for z may be assumed of the form

$$z = k + \frac{m}{a+x} + \frac{n}{b+x}$$

and the values of k, m, n, a, b, determined by weighting the equations in a manner similar to that shown above for the fifth difference curve.

We have used the O^{M} data to illustrate the above process, but generally speaking the latter will be found more useful where the data are only available in large groups, and, in particular, where the limits of the series are not well defined.

In the following table we have a statement taken from Supplement to the Registrar-General's 45th Annual Report, p. cxviii, showing the number of Innkeepers, &c., living at or over certain given ages.

TABLE XVII.

Innkeepers, Publicans, &c. (1881).

Ages t	Living above age t	Proportional numbers $= \frac{1}{\sqrt{\pi}} \int_{\infty}^{z} e^{-t^2} dt$	Values of z
15	232,890	1·0000	∞
20	230,280	·9888	1·6147
25	222,213	·9542	1·1929
45	105,153	·4515	− ·0862
65	14,451	·0620	−1·0877

It will be seen that more than 50 per-cent of the numbers living are in the age-group 25–45, and nearly 40 per-cent in

the group 45–65. In such a series the usual methods of interpolation would probably give unsatisfactory results.

If we treat the values of z as having constant third differences, we obtain the following equations, taking five years of age as the unit—

$$a \qquad\qquad\qquad = 1{\cdot}6147$$

$$a+b \qquad\qquad\quad = 1{\cdot}1929$$

$$a+5b+10c+10d = -{\cdot}0862$$

$$a+9b+36c+84d = -1{\cdot}0877$$

b, c, and d are the values, reckoning from age 20, of the differences of z. The values of a and b are given immediately and solving the remaining equations for c and d we obtain—

$$a = 1{\cdot}6147 \qquad c = {\cdot}04863$$

$$b = -{\cdot}4218 \qquad d = -{\cdot}00782$$

which enable us to form at once the following series c quinquennial age groups.

Table XVIII.

Innkeepers, Publicans, &c. (1881 *Census*).

Age t	Interpolated Values of z	Corresponding Values of $\frac{1}{\sqrt{\pi}}\int_{\infty}^{z} e^{-t^2} - dt$ *	Proportional Population between Age t and $(t+5)$
15	2·0930	·9985	97
20	1·6147	·9888	346
25	1·1929	·9542	774
30	·8197	·8768	1221
35	·4874	·7547	1499
40	·1880	·6048	1533
45	− ·0862	·4515	1376
50	− ·3429	·3139	1119
55	− ·5904	·2020	834
60	− ·8360	·1186	566
65	−1·0876	·0620	34[illegible]
70	−1·3534	·0278	176·7
75	−1·6407	·01017	73·5
80	−1·9577	·00281	22·5
85	− 2·3121	·00053	4·7
90	−2·7117	·00006	·6

Representing the population living above age x out of a total population of 1.

It will be seen that this distribution shows a small number of cases below age 15. This may be avoided if it is desired to commence the curve at and not before that age, by writing

$$z = \frac{m}{t-15} + a + bt + ct^2$$

giving
$$\frac{dz}{dt} = b + 2ct - \frac{m}{(t-15)^2}$$

the term $\frac{m}{t-15}$ being introduced in order to give the high values of z required near the origin, or, we may write as suggested above, in connection with the O^M data,

$$z = k + \frac{m}{a+x} + \frac{n}{b+x}$$

the value of a being taken in this case as equal to -15.

This form for the value of z will be found very convenient where the series is known to be limited in either direction and the number of groups is small. In certain cases either a or b may be known, and we have, then, only four constants, m, n, k, and b or a, to determine, for which four groups will suffice. Or it may be convenient to assume values for both a and b, in which case with four groups we may write

$$x = \frac{m}{t+a} + \frac{n}{t+b} + k + ct,$$

determining m, n, k and c from the data.

In the case of any statistics intended to be used by the actuary, it is important to consider not only how far they are suitable for the purpose for which they are to be employed, but also whether the data are sufficient to render the conclusions drawn from them safe. We have already referred to this question in general terms, but it is necessary to consider it rather more closely.

In practice the actuary has to deal either, (1), with tables based upon a large number of observations; for example, tables such as the O^M, the Government Annuitants, the Manchester Unity Tables of Sickness, &c., where the

accidental errors due to the limited numbers are practically insignificant, but where, on the other hand, there may be uncertainty as to the suitability of the experience for the case in hand; or, (2), with data of more limited extent but known to be applicable, as in the valuation of a pension fund of a Friendly Society by tables based upon its own experience.

In the latter case it is important to be able to form some judgment as to the extent of the probable errors involved in the use of the data and their effect upon the financial values deduced therefrom. This is a problem not susceptible of an exact solution. It is true that if the series of numbers representing the deaths, marriages, or retirements, as the case may be, can be represented by a frequency curve, the probable error of the constants may be obtained in the manner shown by Professor Karl Pearson in his paper on this subject. But these results will be little practical use to us, as the manner in which these probable errors, which are not independent, will affect the monetary values deduced from the graduated rates is too complicated. We can only deal with the problem in a very general manner. We are not even sure that the ordinary theory of errors is applicable to such functions as rates of mortality, sickness, or superannuation; indeed, we may well suspect that it is not strictly applicable.

If the probability of throwing head at a single toss of a coin is one-half, and if in 100 throws 54 heads appear to 46 tails, we do not suppose that the probability of the average number of 50 heads appearing in the next 100 throws is affected. But in the case of the probabilities of death it may well be that an abnormally high or low rate of mortality in a given year may affect the probable rate in succeeding years, and that there may be a tendency for the deviations from the average result to correct themselves, a low rate in a given year leaving a larger number, and a high rate a smaller number, of impaired lives surviving, and thus changing for the time being the constitution of the group under observation.

The "standard deviation" in the value of a_x as deduced from a given experience has not, that I am aware of, been estimated. It will be instructive to attempt this, as an example, for the $O^{M(5)}$ table. It will be sufficient to use approximate methods, as the results will be quite accurate

enough for our purpose. We shall assume that we may take

$$\operatorname{colog}_e p = m = \frac{q}{1-\frac{1}{2}q}$$

and that if the standard deviation in $\log y = \sigma$, then the standard deviation in $y = \sigma y$.*

Taking the observations at a given age x, let us put

exposed to risk	$=n$
graduated or "true" rate of mortality	$=q$
graduated deaths	$=nq \quad =\theta$
actual deaths	$=nq+z=\theta'$
observed value of q	$=q' \quad =q+\frac{z}{n}$,

where, as we have seen, the average value of z is zero, the average value of $z^2 = nq(1-q)$, &c. (*see* p. 110).

Then the observed value of $m = m'$ where

$$m' = \frac{nq+z}{n-\frac{nq+z}{2}} = \frac{q}{1-\frac{q}{2}} + (\text{terms in powers of } z)$$

$$= m + f(z), \text{ say}$$

$$= \operatorname{colog}_e p + f(z)$$

It will be found that the average value of $f(z)$ is not quite zero though very nearly so, being equal to $m\left(\frac{1}{2n}+\frac{1}{4n^2}\right)$ nearly, a quantity that may be neglected; and that the average value of $[f(z)]^2$ is $\frac{m^2}{nq}$ very nearly, and

$$\sqrt{\text{average value of } [f(z)]^2} = \frac{m}{\sqrt{nq}}.$$

Hence, the standard deviation in the "central" death (or marriage or secession or any similar) rate is very nearly equal to the rate divided by the square root of the number of deaths (marriages or secessions, &c.). The errors in $\log_e p$ are of

* If $\log_e y$ have the small error σ, y will be changed to $e^{\log_e y+\sigma} = y.e^{\sigma}$ $= y(1+\sigma+\ldots)$, *i.e.*, the corresponding error in y will be σy nearly.

course the same, but of opposite sign to those in $\text{colog}_e p$. Let the observed value of $\log_e p_x$ be $\log_e p'_x$. We will write

$$\log_e p'_x = \log_e p_x + u_x$$

where u_x is the error of observation whose value in a particular case is fixed but unknown, the average value over a long series of similar observations being zero, and the average value of u^2_x being $\frac{(\text{colog}_e p_x)^2}{nq_x}$ or $\frac{(m_x)^2}{nq_x}$; where nq is the graduated number of deaths at age x.

Taking an arbitrary radix for our mortality table, say l_x, the values of $\log l_{x+t}$ for ages above x will be

$$\log_e l'_x = \log_e l_x$$

$$\log_e l'_{x+1} = \log_e l_{x+1} + u_x$$

$$\log_e l'_{x+t} = \log_e l_{x+t} + (u_x + u_{x+1} + \ldots + u_{x+t-1})$$

similarly, we shall have

$$\log D'_x = \log D_x$$

and for higher ages

$$\log_e D'_{x+t} = \log_e D_{x+t} + (u_x + u_{x+1} + \ldots + u_{x+t-1})$$

whence, on the principle of approximation laid down above,

$$\frac{D'_{x+t}}{D'_x} = \frac{D_{x+t}}{D_x}(1 + u_x + u_{+1} + \ldots u_{x+t-1})$$

Summing this for all values of t from 1 to infinity, we shall have

$$a'_x = \frac{N'_x}{D'_x} = [N_x + (u_x . N_x + u_{x+1} . N_{x+1} + u_{x+2} . N_{x+2} + , \&c.)] \div D_x$$

Here the quantity in the bracket in the numerator is the error in the value of N'_x as deduced from the observations in relation to the value of D'_x corresponding to the arbitrary radix assumed at that age. The average value of each term in the bracket is zero, and the square root of the sum of the average values of the squares of these terms divided by D

will give the standard deviation in the value of a'_x as deduced from the data, which, omitting the suffix x, becomes

$$\frac{1}{\mathrm{D}_0}\sqrt{u_0^2\mathrm{N}_0^2+u_1^2\mathrm{N}_1^2+u_2^2\mathrm{N}_2^2+, \text{\&c.}}$$

If the mortality table be graduated the standard deviation of the graduated values of a'_x will be somewhat less than that of the ungraduated values, but not materially less, except at the ends of the table, the principal effect of the graduation being merely to produce a smooth progression in values.

We might assume, for example, that the effect of graduation was about equivalent to substituting the average error of five successive values of N'_x for the error of the middle value. This would give (omitting a quite insignificant term) the expression

$$\frac{1}{5\mathrm{D}_x}[u_x(3\mathrm{N}_x+\mathrm{N}_{x+1}+\mathrm{N}_{x+2})+u_{x+1}(4\mathrm{N}_{x+1}+\mathrm{N}_{x+2})$$
$$+u_{x+2}.5\mathrm{N}_{x+2}+u_{x+3}.5\mathrm{N}_{x+3}+, \text{\&c.}]$$

for the error in the graduated value of a'_x in lieu of the expression given above.

If we shorten the expression for the standard deviation of a'_x from

$$\frac{1}{\mathrm{D}_x}\sqrt{u_0^2\mathrm{N}_0^2+u_1^2\mathrm{N}_1^2+u_2^2\mathrm{N}_2^2+, \text{\&c.}}$$

to its approximate equivalent

$$\frac{1}{\mathrm{D}_x}\sqrt{5u_2^2.\mathrm{N}_2^2+5u_7^2.\mathrm{N}_7^2+5u_{12}^2.\mathrm{N}_{12}^2+, \text{\&c.},}$$

and, further, take

$$5u_2^2=\frac{25[\mathrm{colog}_e(p_x)]^2}{\text{Observed deaths between } x \text{ and } (x+5)}$$

we shall considerably shorten the labour of calculation, and at the same time, by slightly underestimating the required value, make a rough allowance for the effect of graduation.

We are now in a position to compute a table of standard deviations for a_x for quinquennial intervals of age, the principal steps of the working being set out in the table following. The final columns showing the mean errors or standard deviations in the value of a_x and the corresponding mean errors in P_x found by dividing the former results by the quantity $(1+a_x)^2$.*

*If a_x have an error σ_x, then P_x will have the error

$$\left(\frac{1}{1+a_x}-d\right)-\left(\frac{1}{1+a_x+\sigma_x}-d\right)=\frac{1}{1+a_x}-\frac{1}{1+a_x+\sigma_x}=\frac{\sigma_x}{(1+a_x)(1+a_x+\sigma_x)}$$
$$=\sigma_x\div(1+a_x)^2 \text{ nearly.}$$

TABLE XIX.

Computation of the standard deviations σ_x in the deduced values of a_x and $\sigma_x \div (1+a_x)^2$ in the deduced values of P_x.

Age	$25[\operatorname{colog}_e p_{x+2}]^2 \times 100$	Deaths between Ages x and $x+5$	$5u^2_{x+2}\times 10^4 = 10^2(2)\div(3)$	$5u^2_{x+2}N^2_{x+2} \times 10^{-4}$	Sum of last column	$\frac{10^2\sqrt{\text{col. }(6)}}{D_x} = \sigma_x$	$\frac{\sigma_x}{(1+a_x)^2}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
15	·1020	10	1·020	20244·	21570·	·2200	·00038
20	·1113	122	·0912	1163·	1326·	·0653	·00012
25	·1266	924	·01370	110·0	162·5	·0274	·00006
30	·1525	3,072	·004966	24·48	52·52	·0187	·00004
35	·1981	5,689	·003482	10·20	28·04	·0165	·00004
40	·2813	8,152	·003451	5·758	17·84	·0159	·00005
45	·4410	10,257	·004295	3·864	12·08	·0160	·00006
50	·7632	12,620	·006048	2·726	8·215	0164	·00007
55	1·444	14,903	·009694	1·986	5·489	·0169	·00010
60	2·945	16,618	·01772	1·445	3·503	·0177	·00014
65	6·359	17,455	·03644	·9770	2·059	·0187	·00021
70	14·32	16,042	·08929	·6052	1·082	·0203	·00033
75	33·20	12,172	·2728	·3185	·4764	·0228	·00059
80	78·51	7,317	1·073	·1227	·1580	·0272	·00116
85	188·1	2,865	6·566	·03151	·03528	·0364	·00267
90	454·6	692	65·71	·003659	·003775	·0550	·00705
95	1105·	86	1285·	·000118	·000118	·0966	·02146

The result we have arrived at shows that the mean error, or standard deviation, in the values of the 3 per-cent Annuities in an aggregate experience such as the $O^{M(5)}$ is about one-fiftieth of a year's purchase from about 30 to 65 years of age. Owing to the greater number of deaths at the younger ages in the O^M experience this would about represent standard deviations for that Table from 25 to 65.

If we suppose an experience in which the data were one-hundredth of the extent of the $O^{M(5)}$ but similarly distributed, it is obvious, from a consideration of the process by which the above result was obtained, that the standard deviations or mean errors in the annuity-values would be ten times greater than the values found above. Hence, with an experience including about 1,000 deaths distributed approximately as in the $O^{M(5)}$ data the deduced annuity-values between ages 30 and 60 would on the average be uncertain to about $\pm$·20, or from 1 per-cent to $1\frac{1}{2}$ per-cent of their values. The standard deviations above obtained would be somewhat reduced in a small experience by graduating the experience by Makeham or by a suitable frequency curve, but not very materially. It would occupy too much time

to investigate this point, but we may easily find a limit to the effect of any possible method of graduation in reducing the standard deviations of the annuities. In any ordinary experience such as the O^M, where the observed deaths are a small fraction of the lives passing under observation, the errors in the annuity-values will be due, 1°, to the mortality on the whole being above or below normal, 2°, to the distribution of the mortality being abnormal. This latter factor can alone be affected by any method of graduation. Assume it to disappear altogether, and consider the standard deviation for say a_{50} ($O^{M(5)}$ 3 per-cent) obtained on this hypothesis. There were approximately 100,000 deaths observed above age 50 in this experience. We have $\sqrt{100{,}000}=316$ nearly, and if we assume the mortality above 50 to be throughout subject to an error of $\pm\frac{1}{316}$ of the observed amount, this will be equivalent to changes of $\pm\frac{A}{316}$ and $\pm\frac{B}{316}$ in the values of the constants A and B respectively, which, taking the value of $A=\cdot00589$ and $\log c=\cdot039$, are equivalent in their effect upon the annuity-value to a change of ·00186 in the rate of interest per-cent and of ·0341 years in the age. The combined effect of these changes upon the annuity-value at age 50 is equivalent to $\pm\cdot0148$ as compared with the standard deviation of ·0185 obtained above. The very considerable standard deviations at the ends of the table would, however, be reduced in much greater proportion.

The problem dealt with above is not the same as that of determining the standard deviation in the estimated value of an annuity on a single life. This problem, which is also of importance, has been dealt with by Dr. Bremiker in his paper "On the Risk Attaching to the grant of Life Assurances" (*J.I.A.* xvi, pp. 216, 285). As this paper is not very available for students and the notation is not modern, it may be worth while to give the following short demonstration. For the sake of simplicity "continuous" functions are used.

If the annuitant, aged x at entry, die at the end of the time t the loss to the company granting the annuity, or the deviation from its mean value, referred to the date of entry will be

$$\bar{a}_{\overline{t}|}-\bar{a}_x=\frac{A_x-e^{-t\delta}}{\delta}.$$

and the sum of the squares of all values of this quantity, multiplied by the frequency in each case, will be

$$-\int_0^\infty\left(\frac{\bar{A}_x-e^{-t\delta}}{\delta}\right)^2\frac{d}{dt}\cdot({}_tp_x)\,dt=-\int_0^\infty\frac{(\bar{A}_x)^2-2e^{-t\delta}\bar{A}_x+e^{-2t\delta}}{\delta^2}\cdot\frac{d}{dt}({}_tp_x)\,dt.$$

Noting that

$$-\int_0^\infty\frac{d}{dt}({}_tp_x)\,dt=1$$

$$-\int_0^\infty e^{-t\delta}\frac{d}{dt}({}_tp_x)\,dt=\bar{A}_x$$

and $$-\int_0^\infty e^{-2t\delta}\frac{d}{dt}({}_tp_x)\,dt=\bar{A}'_x \text{ (at rate of interest}=e^{2\delta}-1)$$

we obtain from the above, as the value of the standard deviation of $\bar{a}_x$, and therefore with sufficient accuracy for practical purposes of a_x $(=\bar{a}_x-\frac{1}{2}$ nearly) the expression

$$\sigma=\frac{1}{\delta}\left[\bar{A}'_x-(\bar{A}_x)^2\right]^{\frac{1}{2}}$$

the first term in the bracket being computed at the rate of interest $e^{2\delta}-1$, and the second at the rate $e^{\delta}-1$. It is obvious that the standard deviation for $\bar{A}_x$ will be the above expression multiplied by δ; and for $\bar{A}_x$ less the capitalised value of the annual premiums $(\bar{P}_x)$ (which Dr. Bremiker terms the "Risk attaching to the grant of Life Assurances" by annual premiums) the risk will be the above expression multiplied by $(\bar{P}_x+\delta)$. The premium is here supposed to be payable continuously; if an ordinary annual premium is in question, we should multiply the above expression for σ by (P_x+d). The arithmetical values of these "risks" attaching to grant of assurances or annuities computed at 4 per-cent, according to Heym's mortality table (General Widows Fund of Berlin) are given in the paper referred to, and show, as is obviously the case from general considerations, that the "risk", or average fluctuation whether profit or loss, attaching to the grant of assurances at annual premiums is considerably greater than that attaching to their grant at single premiums.

In practice the important question for a life office, in this connection—and the same considerations apply to other classes of insurance—is the average amount of the annual (or quinquennial) fluctuation in profit due to the deviation of the

death strain from its average or normal amount. In a soundly managed office these fluctuations never approach the point at which stability is remotely threatened, but they become of importance when they are sufficient to produce any serious variation in the rate of Bonus.

The mean square deviation of $\mathring{e}_x$ will be found by putting $\delta=0$ in the expression for σ^2, which in that case takes the indeterminate form $\frac{0}{0}$ which must be evaluated, according to the rules of the Differential Calculus, by differentiating numerator and denominator. The resulting expression takes the same form, so that the process must be repeated, and the limiting value of the expression for σ^2 when $\delta=0$ will be found to be

$$\mathrm{L}_{\cdot\delta=0}\frac{1}{2}\left[\frac{d^2}{d\delta^2}\mathrm{A}'_x-\frac{d^2}{d\delta^2}(\bar{\mathrm{A}}_x)^2\right]$$

which may easily be reduced to the form

$$-\frac{1}{l_x}\int_0 t^2 dl_{x+t}-\left(\frac{1}{l_x}\int_0^\infty t\,dl_{x+t}\right)^2$$

$$=-\frac{1}{l_x}\int_0^\infty t^2 dl_{x+t}-(\mathring{e}_x)^2$$

$$=[\text{mean square duration}-(\text{mean duration})^2]$$

This being the mean square deviation the standard deviation will be

$$\sigma=[\text{mean square duration}-(\text{mean duration})^2]^{\frac{1}{2}}$$

the mean deviation irrespective of sign is approximately $\cdot798\sigma$ and the probable deviation $\cdot674\sigma$, or very nearly $\frac{4}{5}\sigma$ and $\frac{2}{3}\sigma$ respectively.* [*Cf.* De Morgan, Encycl. Metropolitan, Vol. II, p. 460, Art. 149]. If instead of a single risk the average of n risks be taken, all the above quantities will be divided by $\sqrt{n}$.

* The exact values for the mean deviation irrespective of sign of the expectation of life and of the annuity will clearly be ${}_{t|}\mathring{e}_x$ and ${}_{t|}\bar{a}_x$ respectively. Where t is in the first instance equal to $\mathring{e}_x$ and in the second to the term of the continuous annuity certain $\bar{a}_{\overline{t|}}=\bar{a}_x$.

NOTE A.

On the Evaluation of the Successive Moments of the Binomial Expansion of $(p+q)^n$.

These important moments may be found very simply in the following manner. The expanded series being

$$p^n + np^{n-1}q + \frac{n(n-1)}{2}p^{n-2}q^2 + \ldots + npq^{n-1} + q^n$$

$$= u_0 + u_1 + u_2 + \ldots + u_{n-1} + u_n$$

$= \Sigma u_x$, where the subscript is identical with the exponent of q,

the successive moments round the origin will be Σu_x, $\Sigma x u_x$, $\Sigma x^2 u_x$, &c. We will first find the value of Σu_x, $\Sigma x u_x$, $\Sigma x(x-1)u_x$, &c. We have

$$\Sigma u_x = (p+q)^n = 1^n = 1$$

$$\Sigma x u_x = 0 \times p^n + 1 \times np^{n-1}q + 2 \times \frac{n(n-1)}{2}p^{n-2}q^2 + \ldots + nq^n$$

$$= nq\left[p^{n-1} + (n-1)p^{n-2}q + \frac{\overset{(n-1)(n-2)}{\cancel{n(n-1)}}}{2}p^{n-3}q^2 + \ldots + q^{n-1}\right]$$

$$= nq[p+q]^{n-1} = nq.$$

Similarly

$$\Sigma x(x-1)u_x = 1 \times 2 \times \frac{n(n-1)}{2}p^{n-2}q^2 + 2 \times 3 \times \frac{n(n-1)(n-2)}{\lfloor 3}p^{n-3}q^3$$
$$+ \ldots + n(n-1)q^n$$

$$= n(n-1)q^2[p^{n-2} + (n-2)p^{n-3}q + \ldots + q^{n-2}]$$

$$= n(n-1)q^2[p+q]^{n-2} = n(n-1)q^2$$

and similarly we shall find

$$\Sigma x(x-1)(x-2)u_x = n(n-1)(n-2)q^3, \text{ and so on.}$$

Hencê we shall have

$$\Sigma u_x = 1$$

$$\Sigma^2 u_x = \Sigma x u_x = nq$$

$$\Sigma^3 u_x = \Sigma \frac{x(x-1)}{2} u_x = \frac{n(n-1)}{2} q^2$$

$$\Sigma^4 u_x = \Sigma \frac{x(x-1)(x-2)}{6} u_x = \frac{n(n-1)(n-2)}{6} q^3$$

$$\Sigma^5 u_x = \Sigma \frac{x(x-1)(x-2)(x-3)}{24} u_x = \frac{n(n-1)(n-2)(n-3)}{24} q^4$$

by the formulæ on page 59. Hence we have (*see* the demonstration in Note E, page 124), using m_n for the nth moment round the origin

$$\left.\begin{aligned}
m_0 &= \Sigma u_x = 1 \\
m_1 &= \Sigma^2 u_x = nq \\
m_2 &= 2\Sigma^3 u_x + \Sigma^2 u_x = n(n-1)q^2 + nq \\
m_3 &= 6\Sigma^4 u_x + 6\Sigma^3 u_x + \Sigma^2 u_x = n(n-1)(n-2)q^3 \\
&\qquad + 3n(n-1)q^2 + nq \\
m_4 &= 24\Sigma^5 u_x + 36\Sigma^4 u_x + 14\Sigma^3 u_x + \Sigma^2 u_x = n(n-1)(n-2)(n-3)q^4 \\
&\qquad + 6n(n-1)(n-2)q^3 + 7n(n-1)q^2 + nq
\end{aligned}\right\} \ldots \mathrm{A}$$

These last equations may be found directly, by means of successive differentiation, according to a method suggested by Bertrand (*Calcul des Probabilités*, Chap. IV, Art. 62). We have

$$(p+q)^n = p^n + np^{n-1}q + \frac{n.(n-1)}{2} p^{n-2}q^2 + \ldots + npq^{n-1} + q^n$$

$$\frac{d}{dq}(p+q)^n = \Big[0 \times p^n + 1 \times np^{n-1} + 2 \times \frac{n(n-1)}{2} p^{n-2}q + \ldots + (n-1)npq^{n-2} + nq^{n-1}\Big]$$

and $q \cdot \dfrac{d}{dq}(p+q)^n = \Big[1 \times np^{n-1}q + 2 \times \dfrac{n(n-1)}{2} p^{n-2}q^2 + \ldots + nq^n\Big]$

$= \text{1st moment.}$

Similarly, if we differentiate the last series with respect to q, and multiply the result by q (to restore the power of q which is lost in the differentiation) we shall have

$$\left[1^2 \times np^{n-1}q + 2^2 \times \frac{n(n-1)}{2} p^{n-2}q^2 + \ldots + n^2q^n\right]$$

= 2nd moment; and so on, so that

$$[t\text{th moment}] = q\frac{d}{dq}[(t-1)\text{th moment.}]$$

Thus, the first moment

$$= q\frac{d}{dq}(p+q)^n$$

$$= nq(p+q)^{n-1}.$$

Second moment

$$= q\frac{d}{dq}[nq(p+q)^{n-1}] = n(n-1)q^2(p+q)^{n-2} + nq(p+q)^{n-1}$$

Third moment

$$= q\frac{d}{dq}[n(n-1)q^2(p+q)^{n-2} + nq(p+q)^{n-1}]$$

$$= n(n-1)(n-2)q^3(p+q)^{n-3} + 2n(n-1)q^2(p+q)^{n-2}$$
$$+ n(n-1)q^2(p+q)^{n-2} + nq(p+q)^{n-1}$$

$$= n(n-1)(n-2)q^3(p+q)^{n-3} + 3n(n-1)q^2(p+q)^{n-2} + nq(p+q)^{n-1}$$

Fourth moment

$$= q\frac{d}{dq}\ [\text{third moment}]$$

$$= n(n-1)(n-2)(n-3)q^4(p+q)^{n-4} + 3n(n-1)(n-2)q^3(p+q)^{n-3}$$
$$+ 3n(n-1)(n-2)q^3(p+q)^{n-3} + 6n(n-1)q^2(p+q)^{n-2}$$
$$+ n(n-1)q^2(p+q)^{n-2} + nq(p+q)^{n-1}$$

$$= n(n-1)(n-2)(n-3)q^4(p+q)^{n-4} + 6n(n-1)(n-2)q^3(p+q)^{n-3}$$
$$+ 7n(n-1)q^2(p+q)^{n-2} + nq(p+q)^{n-1}$$

Putting unity* for all the powers of $(p+q)$, these expressions are the same as previously found—*see* equations A.

* This may not be done at any earlier stage because the differentiations are with respect to q, *taking p constant*, whereas to substitute $p+q=1$ before finishing the differentiations would make p vary with q.

We have thus obtained the moments round the origin. Thence the moments round the mean may be found by the formulæ on p. 91. Thus

$$\mu_1 = 0$$

$$\mu_2 = m_2 - (m_1)^2 = n(n-1)q^2 + nq - n^2q^2 = nq - nq^2$$

$$= nq(1-q) = npq$$

$$\mu_3 = m_3 - 3m_1 . \mu_2 - m_1{}^3$$

$$= n(n-1)(n-2)q^3 + 3n(n-1)q^2 + nq$$
$$- 3n^2q^2 + 3n^2q^3 - n^3q^3$$

$$= nq - 3nq^2 + 2nq^3 = nq(1 - 3q + 2q^2)$$

$$= nq(1-q)(1-2q) = npq(p-q)$$

$$\mu_4 = m_4 - 4m_1 . \mu_3 - 6m_1{}^2 . \mu_2 - m_1{}^4$$

$$= nq[(n^3 - 6n^2 + 11n - 6)q^3 + 6(n^2 - 3n + 2)q^2 + 7(n-1)q + 1$$
$$- 4nq(1 - 3q + 2q^2) - 6n^2q^2(1-q) - n^3q^3]$$

$$= nq[3(n-2)q^3 - 6(n-2)q^2 + (3n-7)q + 1]$$

which reduces to

$$nq(1-q)[3(n-2)(1-q)q + 1]$$

$$= npq[3(n-2)pq + 1]$$

It is evident that all the even moments must involve p and q symmetrically; while the odd moments will involve a symmetrical function of p and q, together with the factor $(p-q)$, because they must vanish when $p = q$ (*i.e.*, when the curve is symmetrical) and must only change sign when p and q are transposed.

It may be convenient to repeat here the Author's demonstration given, *J.I.A.*, xxvii, 214, of the value of the average deviation from the mean *irrespective of sign*, that is, treating all the deviations as positive.

If we suppose the event to happen m times in the n trials the deviation from the mean number np will be $(m - np)$ which, since $p + q$ is always equal to 1, may be put in the form $[mq - (n-m)p]$.

This will be positive or negative as m is $>$ or $< np$; and the probability of this particular deviation will be

$$\frac{n \ldots (m+1)}{|\underline{n-m}} p^m q^{n-m}$$

The greatest positive deviation will be nq (when the event happens at all the n trials); the greatest negative deviation $-np$ (when it fails at every trial).

Hence, we have the following scheme, in which m is to be taken as the next integer $< np$.

Possible Deviations from Mean Result np.

Sign	Magnitude	Probability	Magnitude × Probability
Positive	Event will happen $nq = n - np$	p^n	np^nq
	$(n-1)q-p$	$np^{n-1}q$	$n(n-1)p^{n-1}q^2 - np^nq$
	$(n-2)q-2p$	$\frac{n.(n-1)}{2}p^{n-2}q^2$	$\frac{n(n-1)(n-2)}{2}p^{n-2}q^3 - n(n-1)p^{n-1}q^2$
	... ...	... ...	... ...
	... ...	... ...	... ...
	$m+1-np =$ $(m+1)q-(n-m-1)p$	$\frac{n \ldots (m+2)}{\lvert\underline{n-m-1}}p^{m+1}q^{n-m-1}$	$\frac{n \ldots (m+1)}{\lvert\underline{n-m-1}}p^{m+1}q^{n-m} - \frac{n \ldots (m+2)}{\lvert\underline{n-m-2}}p^{m+2}q^{n-m-1}$
Negative	$m-np=$ $mq-(n-m)p$	$\frac{n \ldots (m+1)}{\lvert\underline{n-m}}p^mq^{n-m}$	$\frac{n \ldots m}{\lvert\underline{n-m}}p^mq^{n-m+1} - \frac{n \ldots (m+1)}{\lvert\underline{n-m-1}}p^{m+1}q^{n-m}$
	... ...	... ...	... ...
	... ...	... ...	... ...
	$q-(n-1)p$	npq^{n-1}	$npq^n - n.(n-1)p^2q^{n-1}$
	$-np = 0-np$ Event will happen	q^n	$-npq^n$

If the final column of products is examined it will be seen that each positive term is cancelled by a similar negative term in the succeeding product. Hence, the total of the products, that is to say, the average deviation, is zero, showing that np is the true mean result, the positive and negative deviations from which

exactly balance each other. Of the terms above the horizontal line, representing the positive deviations, the sum is, of course, equal to the only uncancelled term, $\dfrac{n\ldots(m+1)}{\lfloor n-m-1}p^{m+1}q^{n-m}$, and similarly of the terms below the line representing the negative deviations the sum is $-\dfrac{n\ldots(m+1)}{\lfloor n-m-1}p^{m+1}q^{n-m}$. Hence, the average magnitude of the deviations, that is, the total of every possible deviation multiplied by its probability, regardless of sign, is

$$2\cdot\frac{n\ldots(m+1)}{\lfloor n-m-1}p^{m+1}q^{n-m}=\frac{2\lfloor n}{\lfloor m\lfloor n-m-1}p^{m+1}q^{n-m}\ \ldots\quad(a)$$

which, since the sum of all the probabilities is necessarily 1, will also be the average or mean deviation. This result is exact, not approximate, but where n and m are large numbers it is necessary to simplify it by the use of Stirling's formula, which gives for large numbers $\lfloor n=\sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n}$ nearly.

Put (a) into the equivalent form

$$\frac{2\lfloor n(n-m)}{\lfloor m\lfloor n-m}p^{m+1}q^{n-m};$$

using Stirling's approximation to the factorials, we have

$$\frac{2}{\sqrt{2\pi}}n^{n+\frac{1}{2}}m^{-(m+\frac{1}{2})}(n-m)^{-(n-m+\frac{1}{2})}(n-m)p^{m+1}q^{n-m}$$

Since m is the integer immediately below np, we may write $m=np-k$; $n-m=nq+k$ (where k is a fraction); hence, we get

$$\sqrt{\frac{2}{\pi}}n^{n+\frac{1}{2}}\left[np\left(1-\frac{k}{np}\right)\right]^{-np-\frac{1}{2}+k}\left[nq\left(1+\frac{k}{nq}\right)\right]^{-nq+\frac{1}{2}-k}p^{np+1-k}q^{nq+k}$$

$$=\sqrt{\frac{2}{\pi}}n^{\frac{1}{2}}p^{\frac{1}{2}}q^{\frac{1}{2}}\left(1-\frac{k}{np}\right)^{-np}\left(1+\frac{k}{nq}\right)^{-nq}\left(\frac{1+\frac{k}{nq}}{1-\frac{k}{np}}\right)^{\frac{1}{2}-k}$$

but where np and nq are large numbers, k being a proper fraction, the last factor is very nearly equal to 1, and $\left(1-\dfrac{k}{np}\right)^{-np}$ and $\left(1+\dfrac{k}{nq}\right)^{-nq}$ are very nearly equal to e^k and e^{-k} respectively; hence, the above expression reduces to

$$\sqrt{\frac{2}{\pi}npq}=\cdot 79788\sqrt{npq}=\frac{4}{5}\sqrt{npq}\text{ very nearly.}$$

Although this result has been obtained on the assumption that np and nq are large, it will be found to be very approximate even for small numbers. As an extreme case, suppose 120 lives at risk, the probability of death in each case being ·02; the "expected" deaths would then be 2·4, and the extent by which the actual deaths would, on the average, exceed or fall short of this number would be given by the formula as

$$\frac{4}{5}\sqrt{2{\cdot}4 \times {\cdot}98} = 1{\cdot}227.$$

The true value of the average deviation given by formula (a) is

$$2 . \frac{120 . 119 . 118}{1{\cdot}2} ({\cdot}02)^3 ({\cdot}98)^{118}$$

$$= 1{\cdot}243,$$

almost identical with the approximate result above.

NOTE B.

On the Use of Logarithms of the Unadjusted Terms of a Series.

Consider the number of cases out of a given series falling into a particular group; or the number of deaths, or analogous events, at a given age or group of ages, accruing out of a given number at risk. Suppose the series to consist of n cases in all, and let the true probability of any case falling into the particular group be p, and let $m=np$. Let the observed number of cases in the group be $m'=m+z$, where, as we have seen, z has an average value of zero, z^2 has an average value of

$$np(1-p)=\frac{m(n-m)}{n};$$

z^3 has an average value of

$$np(1-p)(1-2p)=\frac{m(n-m)(n-2m)}{n^2}, \text{ \&c.}$$

If we operate with the logs of the observed quantities m', we must avoid by arbitrary grouping cases in which m' is zero, or $m':n$ very small when the logs become infinite or very great; but when this is done we shall still find the logs of the ungraduated numbers less on the average than the values of the graduated (or true) numbers. This may be easily seen from a simple example. Let $n=4$, and $p=\frac{1}{2}$, in which case $m=np=2$. The observed values of m' may be anything from 0 to 4, and we shall have the following possible cases:

Values of $m'=m+z$	Relative frequency of these values	log m'	Products (2) × (3)
(1)	(2)	(3)	(4)
0	$\frac{1}{16}$ } $\frac{5}{16}$	$-\infty$ } $-\cdot097$	(say) $-\cdot030$
1	$\frac{4}{16}$	$\cdot000$	
2	$\frac{6}{16}$	$\cdot301$	$\cdot113$
3	$\frac{4}{16}$	$\cdot477$	$\cdot119$
4	$\frac{1}{16}$	$\cdot602$	$\cdot038$
Total	1	...	$\cdot240$

Here, to avoid the cases in which the observed value of m' is zero, we have combined the first two groups, taking four cases in which $m'=1$, for one case in which $m'=0$, thus giving an average value of $m'=\frac{4}{5}$, the logarithm of which is $-\cdot 097$. Notwithstanding this device our average value of $\log m'$ is only $\cdot 240$ as compared with the value of $\log m = \cdot 301$ (where $m=2$ is the true value or average value of m').

Assume that on the average

$$\log [m'(1+k)] = \log m$$

$$= \log [(m+z)(1+k)]$$

$$= \log m + \frac{z}{m} - \frac{z^2}{2m^2} + \frac{z^3}{3m^3}, \text{ \&c.}, + k - \frac{k^2}{2} + \text{ \&c.}$$

Whence

$$k - \frac{k^2}{2} + \text{ \&c.} = -\frac{z}{m} + \frac{z^2}{2m^2} - \frac{z^3}{3m^3}, \text{ \&c.}$$

Insert the average values as given above for z, z^2, &c.,

$$k - \frac{k^2}{2} + \text{\&c.} = \frac{n-m}{2nm} - \frac{(n-m)(n-2m)}{3n^2m^2} + \text{\&c.},$$

or, omitting terms of the second order,

$$k = \frac{n-m}{2nm} \text{ nearly,}$$

which, again omitting terms of the second order, may be written

$$k = \frac{n-m'}{2nm'}$$

$$\log [m'(1+k)] = \log \left[m' + \frac{n-m'}{2n}\right] = \log \left[m' + \frac{1-p'}{2}\right]$$

where p' is the observed value of the probability p.

If this expression be substituted for $\log m'$ in the example given above, we should have as the sum of the products of col. (2) $\times$ col. (3) the value $\cdot 309$, which is very much nearer the true value $\cdot 301$ than the uncorrected value in the above table. If we take larger numbers, as $n=100$, $m=np=10$, we shall find by a similar process the average value of $\log_{10}\left(m' + \frac{1-p'}{2}\right)$ is $\cdot 99987$ as compared with the true value of $\log m = 1\cdot 00000$. Where the numbers n and m' are very large, the correction, of course, becomes insignificant.

It may be shown in like manner that, if we are dealing with the reciprocals of the observed values, then, on the average,

$$\frac{1}{m' + 1 - p'} = \frac{1}{m} \text{ nearly,}$$

and again, on the average,

$$\sqrt{m' + \frac{1 - p'}{4}} = \sqrt{m}$$

Reverting to the question of the use of the logs of the ungraduated quantities, it will be found that if the above results are made use of in practice, the logarithms will be over-corrected. The reason for this is that we do not eventually arrive at the true values of log m and log p, the graduated values being still affected by an outstanding or unbalanced error. If our series consists of a large number of groups, these outstanding errors will be comparatively small, and the above correction will not be much in excess; but if the number of groups is very small, our graduated quantities must necessarily follow rather closely the original values, and the use of the above formula would largely over-correct the series. Suppose, for example, we had a series of ten groups. We should require about five groups to obtain the general form of the curve, or to determine the constants of any frequency curve employed, hence the errors of the groups would only be reduced by the ratio of approximately $\frac{1}{\sqrt{2}}$ and the correction k as shown above should be reduced by half, and proportionately in other cases.

NOTE C.

ON THE RATIONALE OF THE METHOD OF LEAST SQUARES.

IN statistical work it often happens that a number of constants, entering into the known mathematical form of a given function, have to be evaluated from a much greater number of observed values of the function. We may, for example, have three constants, such as x, y, z, in the expression $lx + my + nz = \mathrm{F}$, and fifty observed values of F (embodying different values of the coefficients l, m, n) from which to determine the constants. If the observed values of F were rigidly accurate, any three of them, or any three combinations, would suffice to determine the constants, and it would be immaterial what set of three was selected, since all would lead to the same results. But generally the observed values of F will be affected by errors of observation and hence will not be strictly consistent; and taking the above example each of the $\frac{50 \times 49 \times 48}{6} = 1960$ different sets of three individual equations would in general produce different values of the constants: so that, apart from the prohibitive amount of labour required in the solution of so many equations, we should have no means of deciding which was the best or most advantageous solution, or how to combine the solutions in order to obtain the best average results. The method of least squares supplies the means of combining the original observations in such a manner as to produce a number of equations, equal to the number of unknowns (in the above example, three), the solution of which by the usual process leads to the most probable values of the unknown constants.

Suppose that the observed function F is a linear function of the variables x, y, z, of the form $lx + my + nz \ldots$, and that the errors in the observed values of F follow the "normal law", so that the probability of an error k is proportional to e^{-k^2/c^2}, where the standard deviation of F is $\frac{c}{\sqrt{2}}$. We shall further suppose that the equations

have been so weighted that the value of c is the same in each of the observations, or that the "precision" is uniform. Thus, for example, if in a given equation the probability of an error of k in the observed value of F is proportionate to e^{-k^2/c_1^2}, with a standard deviation of $\frac{c_1}{\sqrt{2}}$, then multiplying the equation by $\frac{c}{c_1}$, we shall have an equation with a standard deviation of $\frac{c}{\sqrt{2}}$ as before, and the probability of an error k will be proportional to e^{-k^2/c^2} as required.

Let there be t equations as follows (where t is supposed greater than the number of unknowns, say s):

$$\left.\begin{array}{l} l_1x+m_1y+n_1z+\ \dots\ -w_1\mathrm{F}=k_1 \\ l_2x+m_2y+n_2z+\ \dots\ -w_2\mathrm{F}=k_2 \\ \vdots \qquad \vdots \qquad \vdots \qquad \vdots \\ l_tx+m_ty+n_tz+\ \dots\ -w_t\mathrm{F}=k_t \end{array}\right\} \quad . \quad . \quad . \quad . \quad . \quad \text{(A)}$$

where F represents the true value of the observed function and $k_1, k_2, \dots$ the errors of observation. The chance of the errors being, by hypothesis, respectively proportional to $e^{-k_1^2/c^2}$, $e^{-k_2^2/c^2}$ ... the chance of the conjunction of these individual errors will be proportional to $e^{-[k_1^2/c^2+k_2^2/c^2+\dots]}$, which will obviously have its greatest value when the quantity in brackets is a minimum. Now, the most probable values of the constants will be those that give the greatest probability of the observed event, *i.e.*, the happening of the given combination of errors. Thus, the most probable values will be those making $[k_1^2/c^2+k_2^2/c^2+\dots]$ or $\Sigma k_t^2/c^2$ or Σk_t^2 a minimum—hence the name "method of least squares."

Now we have

$$\Sigma k_t^2=\Sigma[(l_tx+m_ty+n_tz+\ \dots\ -w_t\mathrm{F})^2]$$

and since x, y, z ... are supposed to be independent, the minimum value must correspond to such values of x, y, z ... as will make the partial differential co-efficients of this expression, with respect to x, y, z ..., all vanish.* Hence we must have, omitting a common factor 2,

$$\left.\begin{array}{l} \Sigma[l_t(l_tx+m_ty+n_tz+\ \dots\ -w_t\mathrm{F})]=0 \\ \Sigma[m_t(l_tx+m_ty+n_tz+\ \dots\ -w_t\mathrm{F})]=0 \\ \Sigma[n_t(l_tx+m_ty+n_tz+\ \dots\ -w_t\mathrm{F})]=0 \\ \quad \&c., \qquad \&c., \qquad \&c. \end{array}\right\} \quad . \quad . \quad . \quad . \quad . \quad \text{(B)}$$

* These conditions, though *necessary*, are not in general *sufficient* to ensure a minimum, but in this case it is obvious that a minimum exists because high *negative* values and high *positive* values of x, y, z ... alike give large values to the function.

(the summation extending to all values of t) as the system of equations, s in number, for determining the most probable values of x, y, z. Hence the rule:

"First prepare the equation by multiplying each by its proper weight (the reciprocal of the probable error or standard deviation), thus giving a set of equations with a uniform p.e. and s.d. Multiply each equation by the coefficient of x and add all the results together; next multiply each by the coefficient of y and add all the results together, and so on: the resulting aggregate equations, solved in the usual manner, give the most probable values of the constants."

It will be seen at once that if there is only one constant to be determined, the method based on the normal law of error gives the weighted average, *i.e.* the total of the weighted values divided by the total weights, as the most probable. Conversely, it may be shown that if the weighted average is the most probable value, then the facility of error must follow the normal law. Apart, however, from any hypothesis as to the law of error, it may be shown mathematically that the method of least squares gives results which become more and more nearly accurate as the number of observations increases. Considerations of a more general kind will also lead to the conclusion that the method must produce very good results. Without giving any definite form to the law of error, it is obvious that large errors are less probable than small, and that the most advantageous system of values for the unknown constants will be that which produces, on the whole, the smallest numerical deviations (irrespective of sign) between the adjusted and observed values of the function. Now, if the law of error is supposed unknown, we cannot investigate mathematically the conditions required to produce a minimum deviation irrespective of sign; and the simplest function of the errors which is independent of sign is the square of the errors, which will be the same for a positive or negative deviation, and at the same time attributes a rapidly increasing importance, or disadvantage, to the errors as they increase in magnitude. Hence we can see, in a very general way, that a method which gives a minimum value to the sum of the squares of the errors, is likely to lead to satisfactory results consistent with elementary notions as to the nature of the errors. Moreover, in actuarial work we usually have to do with numbers sufficiently large to make the normal law of error very near the truth.

Reverting to the system of equations (B), it will easily be seen that if F is a parabolic function of the form $x+ay+a^2z+\ldots$ the equations for determining $x, y, z, \ldots$ &c., are equivalent to reproducing ΣF, ΣaF, Σa^2F (ΣwF, $\Sigma wF.a$, &c., if the equations are weighted), *i.e.*, the successive moments of the observations.

It has, so far, been supposed that the function F is a linear function of the constants $x, y, z, \ldots$ If this is not the case, suppose that the equally weighted equations, from which the values of $x, y, z, \ldots$ are to be found, are of the form

$$\left.\begin{aligned} f_1(x, y, z \ldots) - w_1\mathrm{F} &= k_1 \\ f_2(x, y, z \ldots) - w_2\mathrm{F} &= k_2 \\ \&\text{c.}, \quad \&\text{c.}, \quad &\&\text{c.} \end{aligned}\right\} \quad \ldots \ldots \quad (\mathrm{C})$$

where $f_1, f_2 \ldots$ are known functions of the variables $x, y, z \ldots$ By means of t of these equations, or of t combinations from amongst them, or otherwise, find approximate values of $x, y, z \ldots$ say $x^1, y^1, z^1 \ldots$; and suppose that $x = x^1 + \delta x$, $y = y^1 + \delta y$, $z = z^1 + \delta z$, &c., where it may be supposed that $\delta x, \delta y, \delta z \ldots$, representing small corrections to be found, are so small that their squares may be neglected. Then if

$$f_1 = f_1(x^1, y^1, z^1 \ldots)\,;\ \frac{df_1}{dx^1} = \frac{d}{dx^1} f_1(x^1, y^1, z^1 \ldots)\ \&\text{c.}$$

and so on, equation (C) will become

$$\left.\begin{aligned} f_1 + \delta x\left(\frac{df_1}{dx^1}\right) + \delta y.\left(\frac{df_1}{dy^1}\right) + dz\left(\frac{df_1}{dz^1}\right) + \ldots - w_1\mathrm{F} &= k_1 \\ f_2 + \delta_x\left(\frac{df_2}{dx^1}\right) + \delta y.\left(\frac{df_2}{dy^1}\right) + \delta z\left(\frac{df_2}{dz^1}\right) + \ldots - w_2\mathrm{F} &= k_2 \\ \&\text{c.} \qquad \&\text{c.} \qquad \&\text{c.} & \end{aligned}\right\} \quad (\mathrm{D})$$

These equations are linear functions of the small corrections $\delta x, \delta y, \delta z \ldots$ which can accordingly be found by the rules already derived; and hence are found the corrected values $x = x^1 + \delta x$, $y = y^1 + \delta y$, &c. The process can be repeated, if greater accuracy is desired, until the corrective terms become insignificant.

In the important particular case of a graduation by Makeham's formula, the original equations are of the form

$$w_{x+\frac{1}{2}}.(\mathrm{A} + \mathrm{B}c^{x+\frac{1}{2}}) = w_{x+\frac{1}{2}}\frac{\theta_x}{\mathrm{E}_{x+\frac{1}{2}}} = w_{x+\frac{1}{2}}\,\mu_{x+\frac{1}{2}}$$

(w being the "weight"). Approximate values of the constants, say

A$'$, B$'$, c', being found, the resulting equations for determining δA, δB, δc are as follow, μ'_x representing $A' + B'c'^x$:

$$(\Sigma w)\delta A + (\Sigma w . c'^{x+\frac{1}{2}}.)\delta B + \left(\Sigma w B' c'^{x-\frac{1}{2}} . \overline{x+\frac{1}{2}}\right)\delta c$$

$$= (\Sigma w)(\mu_{x+\frac{1}{2}} - \mu'_{x+\frac{1}{2}})$$

$$(\Sigma w . c'^{x+\frac{1}{2}})\delta A + (\Sigma w c'^{2x+1})\delta B + \left(\Sigma w . B' c'^{2x} . x + \frac{1}{2}\right)\delta c$$

$$= (\Sigma w . c^{x+\frac{1}{2}})(\mu_{x+\frac{1}{2}} - \mu'_{x+\frac{1}{2}})$$

$$(\Sigma w . B'\overline{x + \frac{1}{2}}c'^{x-\frac{1}{2}})\delta A + (\Sigma w . B'\overline{x + \frac{1}{2}}c'^{2x})\delta B$$

$$+ \left\{\Sigma w . B'^2\left(x + \frac{1}{2}\right)^2 c'^{2x-1}\right\}\delta c$$

$$= \Sigma w . B'\overline{x + \frac{1}{2}}c'^{x-\frac{1}{2}}\left(\mu_{x+\frac{1}{2}} - \mu_{x+\frac{1}{2}}\right)$$

For an example, *see J.I.A.*, xvii, 161-71.

NOTE D.

On the Use of the Binomial Curve to Represent a Continuous Series.

If the Binomial curve $y = \dfrac{\underline{|n}}{\underline{|x}\;\underline{|n-x}} p^x q^{n-x}$ made high contact with the axis of x at the points $x = -1$ and $x = n+1$, where y becomes zero, it could be conveniently employed to represent a continuous curve in lieu of representing merely isolated ordinates; as in that case the moments of the continuous curve would very closely agree with those of the isolated ordinates. The same would be true of any series of equidistant points on the curve supposing these to be fairly numerous. If, for example, we suppose the values of y tabulated for every integral value of xh, then the tth moment of the curve would be increased by multiplication by the factor h^t, and from the observed numerical values of the first 4 moments h and the remaining constants could be obtained. As, however, the curve y cuts the axis of x at an angle at both limits, this method of proceeding will lead to approximate results only when n is fairly large.

The area of y treated as a continuous curve may be approximately determined from the well known approximate formula

$$\int_{-1}^{n+1} y.dx = \frac{1}{2} y_{-1} + y_0 + \ldots + y_n + \frac{1}{2} y_{n+1} + \frac{1}{12}\left[\left(\frac{dy}{dx}\right)_{-1} - \left(\frac{dy}{dx}\right)_{n+1}\right]$$

y_{-1} and y_{n+1} being of course equal to zero and the series $y_0 + y_1 + \ldots y_n$ is the expansion of $(p+q)^n$ where we assume $p+q=1$, and is therefore also $=1$. As the factor $\dfrac{1}{\underline{|x}}$ vanishes for $x = -1$; and $\dfrac{1}{\underline{|n-x}}$ vanishes for $x = n+1$, we have

$$\left(\frac{dy}{dx}\right)_{-1} = \left(\frac{\underline{|n}}{\underline{|n-x}} p^x q^{n-x} \frac{d}{dx}\frac{1}{\underline{|x}}\right)_{-1} = \frac{1}{n+1} p^{-1} q^{n+1}$$

$$\left(\frac{dy}{dx}\right)_{n+1} = \left(\frac{\underline{|n}}{\underline{|x}} p^x q^{n-x} \frac{d}{dx}\frac{1}{\underline{|n-x}}\right)_{n+1} = -\frac{1}{n+1} p^{n+1} q^{-1}$$

since $\frac{d}{dx}\frac{1}{\underline{|x}}$ as is known $=1$ when $x=-1$. Hence the area of the curve y becomes

$$\int_{-1}^{n+1} ydx = 1 + \frac{1}{12} \cdot \frac{1}{n+1} \cdot \left(\frac{p^{n+2}+q^{n+2}}{pq}\right)$$

Analogous expressions can be found for the approximate value of the moments

$$\frac{\int xydx}{\int ydx};\ \frac{\int x^2ydx}{\int ydx};\ \&c.,$$

but the relations which result do not lead to sufficiently convenient formulae for practical use.

NOTE E.

On the relations between the Successive Moments and the Successive Summations of a series.

The relations given on p. 60 may be systematically demonstrated, and developed to any extent that may be required, by means of the ordinary interpolation formulæ combined with a table of the power-differences usually known as the "Differences of Nothing" —*see* Text-Book, Part II, Ch. xxii, Art. 11; Sunderland's "Notes on Finite Differences", pp. 24-5.

We have, by the ordinary interpolation formula,

$$v_x = v_0 + x\Delta v_0 + \frac{x.\overline{x-1}}{2}\Delta^2 v_0 + \ldots$$

and hence $u_x . v_x = u_x . v_0 + x u_x . \Delta v_0 + \dfrac{x.\overline{x-1}}{2} u_x . \Delta^2 v_0 + \ldots$

so that $\Sigma(u_x v_x) = (\Sigma u_x) v_0 + (\Sigma x u_x) . \Delta v_0 + \Sigma\left(\dfrac{x.\overline{x-1}}{2} u_x\right) . \Delta^2 v_0 + \ldots$

$$= (\Sigma u_0) v_0 + (\Sigma^2 u_1) . \Delta v_0 + (\Sigma^3 u_2) . \Delta^2 v_0 + \ldots$$

using the notation of p. 60.

Put $\qquad v_x = x^m$ and we have

$$\Sigma x^m . u_x = (\Sigma u_0) 0^m + (\Sigma^2 u_1) \Delta 0^m + (\Sigma^3 u_2) \Delta^2 . 0^m + \ldots$$

Putting m equal successively to 1, 2, 3 . . . , taking the differences from the table of the differences of nothing, and noting that the first term vanishes whatever the value of m, we can write down at once—

$$\left.\begin{aligned}
\Sigma x . u_x &= \Sigma^2 u_1 \\
\Sigma x^2 u_x &= \Sigma^2 u_1 + 2\Sigma^3 u_2 \\
\Sigma x^3 u_x &= \Sigma^2 u_1 + 6\Sigma^3 u_2 + 6\Sigma^4 u_3 \\
\Sigma x^4 u_x &= \Sigma^2 u_1 + 14\Sigma^3 u_2 + 36\Sigma^4 u_3 + 24\Sigma^5 u_4 \\
\Sigma x^5 u_x &= \Sigma^2 u_1 + 30\Sigma^3 u_2 + 150\Sigma^4 u_3 + 240\Sigma^5 u_4 + 120\,\Sigma^6 u_5
\end{aligned}\right\} \;.\,.\; A$$

These equations, divided by Σu_0 give the expressions for the moments set out on page 60.

Taking next the usual central difference formula,

$$v_x = v_0 + xa_0 + \frac{x^2}{\underline{|2}}b_0 + \frac{x(x^2-1)}{\underline{|3}}c_0 + \frac{x^2(x^2-1)}{\underline{|4}}d_0 + \frac{x(x^2-1)(x^2-4)}{\underline{|5}}e_0$$

$$= v_0 + xa_0 + \frac{x^2}{\underline{|2}}b_0 + \frac{(x-1)x(x+1)}{\underline{|3}}c_0 + \frac{x^2(x+1)(x-1)}{\underline{|4}}d_0 + \frac{(x-2)(x-1)x(x+1)(x+2)}{\underline{|5}}e_0$$

$$= v_0 + xa_0 + \frac{x(x-1)+(x+1)x}{2}\cdot\frac{b_0}{\underline{|2}} + (x-1)x(x+1)\frac{c_0}{\underline{|3}} + \frac{\{(x+2)(x+1)x(x-1)\}+\{(x+1)x(x-1)(x-2)\}}{2}\frac{d_0}{\underline{|4}} + \ldots$$

Thus,

$$\Sigma(u_xv_x) = (\Sigma u_x)v_0 + (\Sigma xu_x)a_0 + \frac{1}{2}\{\Sigma x(x-1)u_x + \Sigma(x+1)x.u_x\}\frac{b_0}{\underline{|2}} + \{\Sigma(x+1)x(x-1)u_x\}\frac{c_0}{\underline{|3}} + \ldots$$

$$= (\Sigma u_0)v_0 + \Sigma^2u_1.a_0 + \frac{1}{2}(\Sigma^3u_2 + \Sigma^3u_1)b_0 + \Sigma^4u_2.c_0 + \frac{1}{2}(\Sigma^5u_3 + \Sigma^5u_2)d_0 + \ldots$$

the law of the terms being manifest; or, abbreviating the expression

$$\frac{1}{2}(\Sigma^tu_x + \Sigma^tu_{x+1})$$

by the single symbol $\Sigma^tu_{x+\frac{1}{2}}$, the series may be written

$$\Sigma(u_xv_x) = (\Sigma u_0)v_0 + (\Sigma^2u_1)a_0 + (\Sigma^3u_{1\frac{1}{2}})b_0 + (\Sigma^4u_2)c_0 + \ldots$$

Putting $v_x = x^m$, forming the central differences of x^m as shown in the scheme below, we write down at once

$$\left.\begin{aligned} \Sigma xu_x &= \Sigma^2u_1 \\ \Sigma x^2u_x &= 2\Sigma^3u_{1\frac{1}{2}} \\ \Sigma x^3u_x &= 6\Sigma^4u_2 + \Sigma^2u_1 \\ \Sigma x^4u_x &= 24\Sigma^5u_{2\frac{1}{2}} + 2\Sigma^3u_{1\frac{1}{2}} \end{aligned}\right\} \quad \ldots\ldots\ldots \quad \text{B}$$

$v = x^2$

x	v_x	Δ	Δ^2
-2	4		
		-3	
-1	1		2
		-1	
0	0	(0)	2
		1	
1	1		2
		3	
2	4		

$v = x^3$

x	v_x	Δ	Δ^2	Δ^3
-3	-27			
		19		
-2	-8		-12	
		7		6
-1	-1		-6	
		1		6
0	0	(1)	0	(6)
		1		6
1	1		6	
		7		6
2	8		12	
		19		
3	27			

$v = x^4$

x	v_x	Δ	Δ^2	Δ^3	Δ^4
-3	81				
		-65			
-2	16		50		
		-15		-36	
-1	1		14		24
		-1		-12	
0	0	(0)	2	(0)	24
		1		$+12$	
1	1		14		24
		15		$+36$	
2	16		50		
		65			
3	81				

The simplification in the formulæ is, of course, due to the fact that when m is even the odd central differences vanish, and when m is odd the even central differences vanish.

It is sometimes required to find moments of the form

$$u_{x+\frac{1}{2}} \times \left(x + \frac{1}{2}\right)^m$$

For this purpose we may use the formula (*see* "Sunderland's Notes on Finite Differences," p. 32)—

$$v_x = \tfrac{1}{2}(v_0 + v_1) + (x - \tfrac{1}{2})\Delta v_0 + \frac{x(x-1)}{2}\tfrac{1}{2}\Delta^2(v_0 + v_{-1}) + \frac{x(x-1)(x-\frac{1}{2})}{\underline{|3}}\Delta^3 v_{-1}$$

$$+ \frac{x(x-1)(x+1)(x-2)}{\underline{|4}}\frac{1}{2}\Delta^4(v_{-1} + v_{-2}) +$$

$$= \frac{1}{2}(v_0 + v_1) + \frac{1}{2}(x + x - 1)\Delta v_0 + \frac{x(x-1)}{2}\frac{1}{2}\Delta^2(v_0 + v_{-1})$$

$$+ \frac{1}{2}\cdot\frac{(x+1)x(x-1) + x(x-1)(x-2)}{\underline{|3}}\Delta^3 v_{-1}$$

$$+ \frac{(x+1)x(x-1)(x-2)}{\underline{|4}}\frac{1}{2}\Delta^4(v_{-1} + v_{-2})$$

whence we find, in the same manner as before, that commencing with $v_1 w_1$, we shall have

$$\Sigma v_x . w_x = \Sigma w_1 . \frac{v_0 + v_1}{2} + \Sigma^2 w_{1\frac{1}{2}} . \Delta v_0 + \Sigma^3 w_2 . \frac{1}{2}(\Delta^2 v_0 + \Delta^2 v_{-1})$$

$$+ \Sigma^4 w_{2\frac{1}{2}}\Delta^3 v_{-1} + \Sigma^5 w_3 \frac{1}{2}(\Delta^4 v_{-1} + \Delta^4 v_{-2}) + \ldots$$

Putting $$v_x = \left(-\frac{1}{2}+x\right)^m = \frac{(2x-1)^m}{2^m}$$

the following Table shows the values of $(2x-1)^m$ and its differences, whence v_x and its differences will be found by dividing by 2^m.

x	$2x-1$	$(2x-1)^1$	Δ	$(2x-1)^2$	Δ	Δ^2	$(2x-1)^3$	Δ	Δ^2	Δ^3	$(2x-1)^4$	Δ	Δ^2	Δ^3	Δ^5
-2	-5	-5		25			-125				625				
			2		-16			98				-544			
-1	-3	-3		9		8	-27		-72		81		464		
			2		-8			26		48		-80		-384	
0	-1	-1		1		8	-1		-24		1		80		384
		0	2	(1)	0	(8)	(0)	2	(0)	48	(1)	0	(80)	0	(384)
1	$+1$	$+1$		1		8	$+1$		$+24$		1		80		384
			2		8			26		48		80		384	
2	$+3$	$+3$		9		8	$+27$		$+72$		81		464		
			2		16			98				544			
3	$+5$	$+5$		25			$+125$				625				

Dividing by the appropriate power of 2 and inserting the values of $$\frac{1}{2}(v_0+v_1),\ \Delta v_0,\ \frac{1}{2}(\Delta^2 v_0+\Delta^2 v_{-1}),\ \&c.,$$ the last formula becomes

$$\left.\begin{aligned}\Sigma v_x w_x = \Sigma\left(\frac{2x-1}{2}\right)^m w_x,\ &\text{commencing with } \left(\frac{1}{2}\right)^m w_1\\ =(\text{when } m=1)\ &\Sigma^2 w_{1\frac{1}{2}}\\ =(\text{when } m=2)\ &2\Sigma^3 w_2+\frac{1}{4}\Sigma w_1\\ =(\text{when } m=3)\ &6\Sigma^4 w_{2\frac{1}{2}}+\frac{1}{4}\Sigma^2 w_{1\frac{1}{2}}\\ =(\text{when } m=4)\ &24\Sigma^5 w_3+5\Sigma^3 w_2+\frac{1}{16}\Sigma w_1\end{aligned}\right\} \quad . \quad . \quad (C)$$

Writing now $u_{\frac{1}{2}}=w_0$, and so on, *i.e.*, reckoning the ordinates from zero, so that the moments are of the form $u_{\frac{1}{2}}\left(\frac{1}{2}\right)^m + u_{\frac{3}{2}}\left(\frac{3}{2}\right)^m + \ldots$, these become

$$\left.\begin{aligned}N.m_1 &= \Sigma^2 u_1\\ N.m_2 &= 2\Sigma^3 u_{1\frac{1}{2}}+\frac{1}{4}\Sigma u_{\frac{1}{2}}\\ N.m_3 &= 6\Sigma^4 u_2+\frac{1}{4}\Sigma^2 u_1\\ N.m_4 &= 24\Sigma^5 u_{2\frac{1}{2}}+5\Sigma^3 u_{1\frac{1}{2}}+\frac{1}{16}\Sigma u_{\frac{1}{2}}\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (D)$$

This will be made clearer by a numerical example. Take the following series.

x	u_x	Distance from origin multiplied by 2 $=d$	$u_x \times d$	$u_x \times d^2$	$u_x \times d^3$	$u_x \times d^4$
·5	16·74	1	16·74	16·74	16·74	16·74
1·5	15·69	3	47·07	141·21	423·63	1270·89
2·5	14·70	5	73·50	367·50	1837·50	9187·50
3·5	12·99	7	90·93	636·51	4455·57	31188·99
	60·12		228·24 $\div 2 =$ 114·12	1161·96 $\div 4 =$ 290·49	6733·44 $\div 8 =$ 841·68	41664·12 $\div 16 =$ 2604·01

The alternative method by summation will be as follows:

x	u_x	Σu_x	$\Sigma^2 u_x$	$\Sigma^3 u_x$	$\Sigma^4 u_x$	$\Sigma^5 u_x$
·5	16·74	60·12	144·18 (114·12)	...	...	...
1·5	15·69	43·38	84·06	137·73	204·39 (135·525)	...
2·5	14·70	27·69	40·68	53·67	66·66	79·65
3·5	12·99	12·99	12·99	12·99	12·99	12·99

$$\Sigma^2 u_1 = 114{\cdot}12$$

$$2\Sigma^3 u_{1\frac{1}{2}} + \frac{1}{4}\Sigma u_{\frac{1}{2}} = 2 \times 137{\cdot}73 + \frac{1}{4} \times 60{\cdot}12$$

$$= 275{\cdot}46 + 15{\cdot}03 = 290{\cdot}49$$

$$6\Sigma^4 u_2 + \frac{1}{4}\Sigma^2 u_1 = 6 \times 135{\cdot}525 + \frac{1}{4} \times 114{\cdot}12$$

$$= 813{\cdot}15 + 28{\cdot}53 = 841{\cdot}68$$

$$24\Sigma^5 u_{2\frac{1}{2}} + 5\Sigma^3 u_{1\frac{1}{2}} + \frac{1}{16}\Sigma u_{\frac{1}{2}} = 24 \times 79{\cdot}65 + 5 \times 137{\cdot}73 + \frac{1}{16} \times 60{\cdot}12$$

$$= 1911{\cdot}60 + 688{\cdot}65 + 3{\cdot}76 = 2604{\cdot}01$$

With a heavy series of terms, the saving of labour by the summation method will, as may easily be seen, be very considerable. A further saving of labour may be obtained by calculating the moments round some convenient central point, and thus breaking up the series into two parts in the manner indicated in Mr. Elderton's treatise, pp. 22-33; and any of the formulæ described in these notes may be applied in this manner.

NOTE F

On the Identity of the Method of Moments and Method of Least Squares in the Case of an Exponential Function.

Suppose y an exponential function of x so that

$$y = e^{a+bx+cx^2+\&c.} = e^z, \text{ say.}$$

Then if y be taken to represent any group in a frequency distribution where the number of groups is large the probable error in y will be approximately $\sqrt{y}$. Assume the true values of y, *i.e.*, the true values of a, b, c . . . , to be approximately known, and let the observed values of y be denoted by y'. If, then, we weight each equation $y' - e^z = 0$ by the factor $\frac{1}{\sqrt{y}}$, writing

$$\frac{1}{\sqrt{y}}(y' - e^z) = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

we shall have a series of equations of condition in which the probable error is in each case identical; that is to say, they will be suitably weighted for the application of the method of least squares (*see* Note C., p. 117-8).

Writing $$y^1 = (y + \delta a.\frac{dy}{da} + \delta b.\frac{dy}{db} + \&c.)$$

$$= y\ (1 + \delta a + x.\delta b + x^2.\delta c, \&c.)$$

equation (1) becomes

$$\frac{1}{\sqrt{y}}[y(1 + \delta a + x.\delta b + \&c.) - e^z] = 0 \quad . \quad . \quad . \quad . \quad (2)$$

and multiplying each equation successively by the coefficients of δa, δb, &c., *i.e.*, by $\frac{y}{\sqrt{y}}$, $\frac{y}{\sqrt{y}}x$, $\frac{y}{\sqrt{y}}x^2$, &c., and taking the sum of each

set of products, according to the rules of the method of least squares. we get

$$\Sigma[y(1+\delta a+x.\delta b+x^2.\delta c, \&c.)-e^z]=0$$

$$\Sigma[yx(1+\delta a+x.\delta b+x^2.\delta c, \&c.)-x.e^z]=0$$

&c., &c.

as the system of equations for determining, according to the method of least squares, the small corrections to be applied to the approximate values $a, b, c, \ldots$ used in obtaining the approximate values of y.

Now, obviously, if y is so taken that

$$\Sigma(y-e^z) \quad =0$$

$$\Sigma x(y-e^z) =0$$

$$\Sigma x^2(y-e^z)=0, \ \&c.$$

i.e., if the values of the constants a, b, c are found by the method of moments, &c., the above equations are satisfied by $\delta a=\delta b=\delta c=0$; that is to say, the corrections are zero, or the values found for $a, b, c \ldots$ by the method of moments are in conformity with the method of least squares on the assumption that the observations are properly weighted by multiplying by the factors $\frac{1}{\sqrt{y}}$, the weights being assumed invariable. It may, however, be supposed that small variations in the constants, $a, b, c, \ldots$ would produce slight variations in the weights, in which case other solutions may exist which would also lead, by the method of least squares, to equations satisfied by $\delta a=\delta b=\delta c=0$; but as it is well known that small differences in weights have practically no effect on the results, it is evident that any such alternative solution must be very close to that already formed.

NOTE G.

On obtaining the value of Makeham's constant c direct from the exposures and deaths.

As stated in the text an exact value of this constant is not very important, and this may be illustrated by reference to the data for ascending premium assurances given in Table X. An approximate value for c may readily be found by a process such as the following, which is in principle analogous to the aggregate method employed by Mr. King in the Text-Book, Part II. Take the values of μ for the central age of each group in Table X. Reject the initial and final values, as depending upon only two and three deaths respectively. Take the six values for central ages $32\frac{1}{2}$ to $57\frac{1}{2}$, weighted respectively by the factors 1, 3, 5, 5, 3, 1; weight the six values for central ages $47\frac{1}{2}$ to $72\frac{1}{2}$, and also for $62\frac{1}{2}$ to $87\frac{1}{2}$ in the same manner. We shall then have the following totals:

$\mu_{32\frac{1}{2}} \times 1 = \cdot 0119$	$\mu_{47\frac{1}{2}} \times 1 = \cdot 0137$	$\mu_{62\frac{1}{2}} \times 1 = \cdot 0340$
$\mu_{37\frac{1}{2}} \times 3 = \cdot 0345$	$\mu_{52\frac{1}{2}} \times 3 = \cdot 0534$	$\mu_{67\frac{1}{2}} \times 3 = \cdot 1647$
$\mu_{42\frac{1}{2}} \times 5 = \cdot 0655$	$\mu_{57\frac{1}{2}} \times 5 = \cdot 1160$	$\mu_{72\frac{1}{2}} \times 5 = \cdot 3660$
$\mu_{47\frac{1}{2}} \times 5 = \cdot 0685$	$\mu_{62\frac{1}{2}} \times 5 = \cdot 1700$	$\mu_{77\frac{1}{2}} \times 5 = \cdot 5720$
$\mu_{52\frac{1}{2}} \times 3 = \cdot 0534$	$\mu_{67\frac{1}{2}} \times 3 = \cdot 1647$	$\mu_{82\frac{1}{2}} \times 3 = \cdot 7540$
$\mu_{57\frac{1}{2}} \times 1 = \cdot 0232$	$\mu_{72\frac{1}{2}} \times 1 = \cdot 0732$	$\mu_{87\frac{1}{2}} \times 1 = \cdot 3379$
$S_1 = \cdot 2570$	$S_2 = \cdot 5910$	$S_3 = 2 \cdot 1286$

If the mortality follows Makeham's law, we shall have

$$\frac{S_3 - S_2}{S_2 - S_1} = \frac{1 \cdot 5376}{\cdot 3340} = c^{15}$$

since 15 years is the interval between the centres of our empirical groups. This gives $\log c = \cdot 0442$ nearly. If we take the sum of

the unweighted values of μ in three groups for ages $32\frac{1}{2}$ to $47\frac{1}{2}$, $52\frac{1}{2}$ to $67\frac{1}{2}$, and $72\frac{1}{2}$ to $87\frac{1}{2}$, we should obtain in similar manner.

$$c^{20} = \frac{S_3 - S_2}{S_2 - S_1} = \frac{\cdot 6136}{\cdot 0797} \text{ giving } \log c = \cdot 0443.$$

We may conclude, therefore, that log c probably lies between ·044 and ·045. The values of μ for ages $27\frac{1}{2}$ and $92\frac{1}{2}$, which we have omitted in the foregoing, are respectively much below and much above the general curve. If these values had been included duly weighted, we should have obtained a slightly larger value of log c, nearer to ·045.

If we adopt ·045 as an approximate value, we obtain for the values of the constants A and B, by the process described on p. 65,

$$A = \cdot 00950 \qquad B = \cdot 00003712$$

We will call this curve (α), the deviations from the adjusted values of θ in Table X being shown in the Table below. We might

Ascending Premium Assurance Experience.

Deviations in Computed Deaths for Curves (α) *and* (β).

Middle Age of Group	Observed Deaths corrected as per Table (X)	Deviations: Computed Deaths — Observed Deaths: Curve (α) log c=·045		Curve (β) log c=·046	
		+	−	+	−
$27\frac{1}{2}$	·8	·9	...	1·0	...
$32\frac{1}{2}$	29·2	...	3·5	...	2·8
$37\frac{1}{2}$	102·0	...	1·8	...	·2
$42\frac{1}{2}$	175·2	...	7·3	...	5·9
$47\frac{1}{2}$	191·7	12·8	...	13·0	...
$52\frac{1}{2}$	218·6	3·3	...	2·0	...
$57\frac{1}{2}$	228·4	7·3	...	5·1	...
$62\frac{1}{2}$	255·4	...	2·7	...	5·2
$67\frac{1}{2}$	274·4	...	24·8	...	26·5
$72\frac{1}{2}$	205·6	12·0	...	12·0	...
$77\frac{1}{2}$	151·5	12·1	...	13·6	...
$82\frac{1}{2}$	84·8	...	6·6	...	5·1
$87\frac{1}{2}$	22·3	...	·5	·2	...
$92\frac{1}{2}$	2·1	...	1·1	...	1·1
	Sum of deviations	48·4	48·3	46·9	46·8
	Second sum . .	38·9	37·5	39·3	39·9
	Third sum . .	13·7	71·8	31·9	36·9

expect from our first rough approximation to log c that a smaller value than ·045, say ·044, would give better results. We find, however, that the third sum of the errors of the (α) curve is negative, and this indicates an increase in the value of log c.

Since a higher value of c hollows out the curve at the middle ages, increasing the computed deaths at the extremes of the table, it is clear that the effect must be to increase the third sum of the graduated deaths.

The probalility is therefore, that curve (α) will not be much improved by changing the value of c.

If we take the alternative value log $c = \cdot 046$ we find the deviations from the adjusted values of θ in Table X are given for curve β on the previous page.

There is little to choose between the two graduations, notwithstanding the smallness of the third sum of the deviations in curve (β), for against this may be put the fact that the three largest errors in (α) are all increased in (β). On the whole the curves may be taken as showing that an approximate value of log c is generally sufficient, and that nothing is gained by computing this constant to several places of decimals.

It may at first sight appear inconsistent with the general theory to adopt values of the three constants which do not make the third sum vanish ; *i.e.*, the third moment of the graduated and ungraduated figures identical. It must, however, be remembered that the method of least squares (and with it the method of moments) assumes that the form of the curve is known *a priori*, in which case the method gives the means of determining the most probable values of the constants involved. When, however, we are dealing with a mortality experience, we have no *a priori* right to assume that Makeham's law is strictly applicable; and, if it is not, the deviations instead of following the normal law as assumed in the theory of least squares, will include systematic deviations due to departure from the Makeham law. In these circumstances the method of least squares is not strictly applicable, and we are therefore justified in allowing other considerations to guide us in selection of the constants.

We may here note that if the exposures are represented by a frequency curve, the deaths being recomputed to correspond to the graduated exposures, then the value of log c may, in general, be calculated from the moments of the exposures and of the recomputed deaths. This can readily be done if the exposures are represented

by a binomial curve (*see* Calderon, *J.I.A.*, xxxv, 157), although precautions must be taken so to group data that the number of terms in the binomial is not great—not more, say, than five or six; or by the normal frequency curve (*see* Elderton's "Frequency Curves", pp. 98–100); or by the curve $y = kx^m e^{-\gamma x}$, where, if E_0, E_1, &c., represent the successive moments for the exposures round the origin, and θ_0, θ_1, &c., the similar moments of the recomputed deaths,

we shall have
$$\frac{\gamma - \log_e c}{\gamma} = \frac{\left(\frac{\theta_1}{E_1} - \frac{\theta_0}{E_0}\right)}{\left(\frac{\theta_2}{E_2} - \frac{\theta_1}{E_1}\right)}$$

whence, γ being known, $\log_e c$ is easily found.

The above relation may be thus demonstrated. The force of mortality at age x is assumed to be of the form $A + Bc^x = A + Be^{x \log c} = A + Be^{x\lambda}$, putting $\lambda = \log_e c$. Thus the death curve will be of the form $A . kx^m e^{-\gamma x} + B . kx^m e^{-(\gamma - \lambda)x}$, where the second term is of the same form as the first with $\gamma - \lambda$ substituted for γ. But by the well-known properties of the Gamma integral (*see* Williamson's "Integral Calculus", Art. 120) we have

$$\int_0^\infty e^{-xz} x^m dx = \frac{m}{z} \int_0^\infty e^{-xz} x^{m-1} . dx$$

whence it is easily seen that, writing E'_0, E'_1 . . . for the moments of $kx^m e^{-(\gamma - \lambda)x}$,

$$E_0 = E_0 \qquad \theta_0 = AE_0 + BE'_0$$

$$E_1 = E_0 \times \frac{m+1}{\gamma} \qquad \theta_1 = AE_1 + BE'_0 \frac{m+1}{\gamma - \lambda}$$

$$E_2 = E_0 \times \frac{(m+2)(m+1)}{\gamma^2} \qquad \theta_2 = AE_2 + BE'_0 \frac{(m+2)(m+1)}{(\gamma - \lambda)^2}$$

whence

$$\theta_0 \div E_0 = A + B\frac{E'_0}{E_0} \qquad = A + B'$$

$$\Delta = B' \frac{\lambda}{\gamma - \lambda}$$

$$\theta_1 \div E_1 = A + B\frac{E'_0}{E_0}\frac{\gamma}{\gamma - \lambda} \qquad = A + B'\frac{\gamma}{\gamma - \lambda}$$

$$\Delta_1 = B' \frac{\gamma\lambda}{(\gamma - \lambda)^2}$$

$$\theta_2 \div E_2 = A + B\frac{E'_0}{E_0}\frac{\gamma^2}{(\gamma - \lambda)^2} = A + B'\left(\frac{\gamma}{\gamma - \lambda}\right)^2$$

so that
$$\Delta \div \Delta_1 = \frac{\gamma - \lambda}{\gamma} = \frac{\gamma - \log_e c}{\gamma}$$

If the exposures, as often happens, can only be represented by a curve of the form $y=kx^{\alpha}(1-x)^{\beta}$ (where x represents a proportionate part of the range of the curve so that x ranges between 0 and 1) and if, as before, we represent the successive moments for exposures and deaths by m_0, m_1, &c., M_0, M_1, &c., where m_0 and M_0 are made $=1$, then writing

$$(\alpha+1)-(\alpha+\beta+2)M_1=R_0$$

$$(\alpha+1)M_1-(\alpha+\beta+2)M_2=R_1$$

it will be found that, putting r for the range of the curve in years of age,

$$r\log_e c=\frac{R_0}{(M_2-M_1)-h(m_2-m_1)}$$

$$=\frac{R_1}{(M_3-M_2)-h(m_3-m_2)}$$

from which as the numerical value of all the quantities except $\log_e c$ and h are known, these two may be easily found.

This may be shown as follows :—

Let the curve of exposed to risk be represented by the type

$$y=kx^{\alpha}(1-x)^{\beta}$$

where the entire range of the curve is taken as unity, and assume k, α, and β to be determined in the usual manner.

Let the curve of the recomputed deaths be of the form

$$Akx^{\alpha}(1-x)^{\beta}+Bkx^{\alpha}(1-x)^{\beta}e^{\gamma x}=Ay+Bz \quad . \quad . \quad . \quad (1)$$

i.e., we assume that $-\dfrac{d}{dx}(\log l_x)=A+Be^{\gamma x}$

As regards the curve z, we shall have

$$\log z=\alpha\log x+\beta\log(1-x)+\gamma x$$

$$\frac{dz}{dx}=\left(\frac{\alpha}{x}-\frac{\beta}{1-x}+\gamma\right)z$$

or, multiplying both sides by $x^{t+1}(1-x)$,

$$(x^{t+1}-x^{t+2})\frac{dz}{dx}=[\alpha x^{t}-(\alpha+\beta)x^{t+1}+\gamma(x^{t+1}-x^{t+2})]z \quad . \quad . \quad (2)$$

Integrating the left-hand side of this equation by parts, and noting that the factor $(x^{t+1}-x^{t+2})$ is zero for the limits 1 and 0,

$$\int_0^1[(t+2)x^{t+1}-(t+1)x^t]zdx=\int_0^1[\alpha x^t-(\alpha+\beta)x^{t+1}+\gamma(x^{t+1}-x^{t+2})]zdx$$

that is

$$(t+2)m'_{t+1}-(t+1)m'_t=\alpha m'_t-(\alpha+\beta)m'_{t+1}+\gamma(m'_{t+1}-m'_{t+2})$$

and

$$(\alpha+t+1)m'_t-(\alpha+\beta+t+2)m'_{t+1}+\gamma(m'_{t+1}-m'_{t+2})=0 \quad . \quad . \ (3)$$

there m'_t represents the tth moment of the curve z round the ordinate $x=0$.

If $\gamma=0$, the curve z becomes identical with y, and writing m_t for the tth moment of y round the ordinate $x=0$, we have

$$(\alpha+t+1)m_t-(\alpha+\beta+t+2)m_{t+1}=0 \quad . \quad . \quad . \quad . \quad (4)$$

Write, as before, the total of the exposed $=E_0$, and of the deaths $=\theta_0$, respectively, and represent the total of the exposed multiplied at each age by the factor $e^{\gamma x}$ by E'_0.

Let E_t and θ_t be the tth moments of the curve of exposed to risk and of the recomputed deaths, the areas of the curves not being taken as $=1$, but having the values E_0 and θ_0 above defined, that is to say, representing the total exposures and the total deaths. And let E'_t be the tth moment of the curve of exposures multiplied at each age by $e^{\gamma x}$.

Then we have $\qquad \theta_t=AE_t+BE'_t \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$

where θ_t and E_t are known, but the remaining quantities unknown.

From (3) and (4)

$$(\alpha+t+1)E_t-(\alpha+\beta+t+2)E_{t+1}=0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

and $\quad (\alpha+t+1)E'_t-(\alpha+\beta+t+2)E'_{t+1}+\gamma(E'_{t+1}-E'_{t+2})=0 \quad . \quad (7)$

Write $\qquad (\alpha+t+1)\theta_t-(\alpha+\beta+t+2)\theta_{t+1}=R_t,$

from (5)

$$(\alpha+t+1)(AE_t+BE'_t)-(\alpha+\beta+t+2)(AE_{t+1}+BE'_{t+1})=R_t$$

and from (6)

$$(\alpha+t+1)BE'_t-(\alpha+\beta+t+2)BE'_{t+1}=R_t$$

and from (7)

$$B\gamma(E'_{t+2}-E'_{t+1})=R_t \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Since from (5)

$$BE'_t = \theta_t - AE_t$$

we have $$\gamma[(\theta_{t+2} - \theta_{t+1}) - A(E_{t+2} - E_{t+1})] = R_t \quad . \quad . \quad . \quad . \quad (9)$$

writing $t = 0$ and $t = 1$ respectively, we get

$$[(\theta_2 - \theta_1) - A(E_2 - E_1)] = R_0$$

$$\gamma[(\theta_3 - \theta_2) - A(E_3 - E_2)] = R_1$$

whence $$R_1[(\theta_2 - \theta_1) - A(E_2 - E_1)] = R_0[(\theta_3 - \theta_2) - A(E_3 - E_2)]$$

and $$A = \frac{R_1(\theta_2 - \theta_1) - R_0(\theta_3 - \theta_2)}{R_1(E_2 - E_1) - R_0(E_3 - E_2)} \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

also, from (9) $$\gamma = \frac{R_0}{(\theta_2 - \theta_1) - A(E_2 - E_1)}$$

The value of B cannot be determined directly from these equations as it enters symmetrically with the values of E'_t. It is therefore necessary, having found the value of γ, to compute the value of E'_0 and thence deduce B from equation (5).

Unless the mortality follows Makeham's law very closely better results will be obtained by calculating both E'_0 and E'_1 and obtaining values of A and B satisfying the equations

$$\left.\begin{aligned} AE_0 + BE'_0 &= \theta_0 \\ AE_1 + BE'_1 &= \theta_1 \end{aligned}\right\}$$

Tables of Values of $y = \frac{2}{\sqrt{\pi}} \int_0^z e^{-x^2} . dx.$

	y	Δ	z	y	Δ	z	y	Δ
·01	·01128	1128	·51	·52924	866	1·01	·84681	403
·02	·02256	1128	·52	·53790	856	1·02	·85084	394
·03	·03384	1127	·53	·54646	848	1·03	·85478	387
·04	·04511	1126	·54	·55494	838	1·04	·85865	379
·05	·05637	1125	·55	·56332	830	1·05	·86244	370
·06	·06762	1124	·56	·57162	820	1·06	·86614	363
·07	·07886	1122	·57	·57982	810	1·07	·86977	356
·08	·09008	1120	·58	·58792	802	1·08	·87333	347
·09	·10128	1118	·59	·59594	792	1·09	·87680	341
·10	·11246	1116	·60	·60386	782	1·10	·88021	332
·11	·12362	1114	·61	·61168	773	1·11	·88353	326
·12	·13476	1111	·62	·61941	764	1·12	·88679	318
·13	·14587	1108	·63	·62705	754	1·13	·88997	311
·14	·15695	1105	·64	·63459	744	1·14	·89308	304
·15	·16800	1101	·65	·64203	735	1·15	·89612	298
·16	·17901	1098	·66	·64938	725	1·16	·89910	290
·17	·18999	1094	·67	·65663	715	1·17	·90200	284
·18	·20093	1091	·68	·66378	706	1·18	·90484	278
·19	·21184	1086	·69	·67084	696	1·19	·90761	270
·20	·22270	1082	·70	·67780	687	1·20	·91031	265
·21	·23352	1078	·71	·68467	676	1·21	·91296	257
·22	·24430	1072	·72	·69143	667	1·22	·91553	252
·23	·25502	1068	·73	·69810	658	1·23	·91805	246
·24	·26570	1063	·74	·70468	648	1·24	·92051	239
·25	·27633	1057	·75	·71116	638	1·25	·92290	234
·26	·28690	1052	·76	·71754	629	1·26	·92524	227
·27	·29742	1046	·77	·72382	619	1·27	·92751	222
·28	·30788	1040	·78	·73001	609	1·28	·92973	217
·29	·31828	1035	·79	·73610	600	1·29	·93190	211
·30	·32863	1028	·80	·74210	590	1·30	·93401	205
·31	·33891	1022	·81	·74800	581	1·31	·93606	201
·32	·34913	1015	·82	·75381	571	1·32	·93807	195
·33	·35928	1008	·83	·75952	562	1·33	·94002	189
·34	·36936	1002	·84	·76514	553	1·34	·94191	185
·35	·37938	995	·85	·77067	543	1·35	·94376	180
·36	·38933	988	·86	·77610	534	1·36	·94556	175
·37	·39921	980	·87	·78144	525	1·37	·94731	171
·38	·40901	973	·88	·78669	515	1·38	·94902	165
·39	·41874	965	·89	·79184	507	1·39	·95067	162
·40	·42839	958	·90	·79691	497	1·40	·95229	156
·41	·43797	950	·91	·80188	489	1·41	·95385	153
·42	·44747	942	·92	·80677	479	1·42	·95538	148
·43	·45689	934	·93	·81156	471	1·43	·95686	144
·44	·46623	925	·94	·81627	462	1·44	·95830	140
·45	·47548	918	·95	·82089	453	1·45	·95970	135
·46	·48466	909	·96	·82542	445	1·46	·96105	132
·47	·49375	900	·97	·82987	436	1·47	·96237	128
·48	·50275	892	·98	·83423	428	1·48	·96365	125
·49	·51167	883	·99	·83851	419	1·49	·96490	121
·50	·52050	874	1·00	·84270	411	1·50	·96611	117

Table of Values of $y=\frac{2}{\sqrt{\pi}}\int_0^z e^{-x^2}.dx$—*continued.*

z	y	Δ	z	y	Δ	z	y	Δ
1·51	·96728	117	2·01	·995525	195	2·51	·9996143	202
1·52	·96841	111	2·02	·995720	186	2·52	·9996345	192
1·53	·96952	107	2·03	·995906	180	2·53	·9996537	183
1·54	·97059	103	2·04	·996086	172	2·54	·9996720	173
1·55	·97162	101	2·05	·996258	165	2·55	·9996893	165
1·56	·97263	97	2·06	·996423	159	2·56	·9997058	157
1·57	·97360	95	2·07	·996582	152	2·57	·9997215	149
1·58	·97455	91	2·08	·996734	146	2·58	·9997364	141
1·59	·97546	89	2·09	·996880	141	2·59	·9997505	135
1·60	·97635	86	2·10	·997021	134	2·60	·9997640	127
1·61	·97721	83	2·11	·997155	129	2·61	·9997767	121
1·62	·97804	80	2·12	·997284	123	2·62	·9997888	115
1·63	·97884	78	2·13	·997407	118	2·63	·9998003	109
1·64	·97962	76	2·14	·997525	114	3·64	·9998112	103
1·65	·98038	72	2·15	·997639	108	2·65	·9998215	98
1·66	·98110	71	2·16	·997747	104	2·66	·9998313	93
1·67	·98181	68	2·17	·997851	100	2·67	·9998406	88
1·68	·98249	66	2·18	·997951	95	2·68	·9998494	84
1·69	·98315	64	2·19	·998046	91	2·69	·9998578	79
1·70	·98379	62	2·20	·998137	87	2·70	·9998657	75
1·71	·98441	59	2·21	·998224	84	2·71	·9998732	71
1·72	·98500	58	2·22	·998308	80	2·72	·9998803	67
1·73	·98558	55	2·23	·998388	76	2·73	·9998870	63
1·74	·98613	54	2·24	·998464	73	2·74	·9998933	61
1·75	·98667	52	2·25	·998537	70	2·75	·9998994	57
1·76	·98719	50	2·26	·998607	67	2·76	·9999051	54
1·77	·98769	48	2·27	·998674	64	2·77	·9999105	51
1·78	·98817	47	2·28	·998738	61	2·78	·9999156	48
1·79	·98864	45	2·29	·998799	58	2·79	·9999204	46
1·80	·98909	43	2·30	·998857	55	2·80	·9999250	43
1·81	·98952	42	2·31	·998912	53	2·81	·9999293	41
1·82	·98994	41	2·32	·998965	51	2·82	·9999334	38
1·83	·99035	39	2·33	·999016	49	2·83	·9999372	37
1·84	·99074	37	2·34	·999065	46	2·84	·9999409	34
1·85	·99111	36	2·35	·999111	44	2·85	·9999443	33
1·86	·99147	35	2·36	·999155	42	2·86	·9999476	31
1·87	·99182	34	2·37	·999197	40	2·87	·9999507	29
1·88	·99216	32	2·38	·999237	38	2·88	·9999536	27
1·89	·99248	31	2·39	·999275	36	2·89	·9999563	26
1·90	·99279	30	2·40	·999311	35	2·90	·9999589	109
1·91	·99309	29	2·41	·999346	33	2·95	·9999698	81
1·92	·99338	28	2·42	·999379	32	3·00	·9999779	60
1·93	·99366	26	2·43	·999411	30	3·05	·9999839	45
1·94	·99392	26	2·44	·999441	28	3·10	·9999884	32
1·95	·99418	25	2·45	·999469	28	3·15	·9999916	24
1·96	·99443	23	2·46	·999497	26	3·20	·9999940	29
1·97	·99466	23	2·47	·999523	24	3·30	·9999969	16
1·98	·99489	22	2·48	·999547	24	3·40	·9999985	8
1·99	·99511	21	2·49	·999571	22	3·50	·9999993	3
2·00	·99532	20	2·50	·999593	21	3·60	·9999996	

TABLE OF

[The constants are restricted to positive quantities of significant value

Type	Character of Curve: Shape	Character of Curve: Range	Equation $y=$	Limits of x: Lower	Limits of x: Upper	Mean $=\mu_1$	μ_2	μ_3
I	Symmetrical	Limited both ways	$k(a^2-x^2)^{\frac{n}{2}-1}$	$-a$	$+a$	0	$\frac{a^2}{n+1}$	0
II	Symmetrical	Un-limited	$ke^{-\frac{x^2}{a^2}}$ (Normal Curve)	$-\infty$	$+\infty$	0	$\frac{1}{2}a^2$	0
III	Symmetrical	Un-limited	$k(a^2+x^2)^{-\left(\frac{n}{2}+1\right)}$ $(n>3)$	$-\infty$	$+\infty$	0	$\frac{a^2}{n-1}$	0
IV	Skew	Limited both ways	$k(a-x)^{np-1}(a+x)^{nq-1}$ $(p+q=1)$	$-a$	$+a$	$(q-p)a$	$\frac{4pq}{n+1}a^2$	$\frac{16(p-q)pq}{(n+1)(n+2)}$
V	Skew	Limited one way	$kx^{m-1}e^{-\frac{x}{a}}$	0	$+\infty$	ma	ma^2	$2ma^3$
VI	Skew	Limited one way	$k(x-a)^{np-1}(x+a)^{-(nq+1)}$ $(q-p=1)$ $(n>3)$	$+a$	$+\infty$	$(p+q)a$	$\frac{4pq}{n-1}a^2$	$\frac{16(p+q)pq}{(n-1)(n-2)}a^3$
VII	Skew	Limited one way	$kx^{-(n+2)}e^{-\frac{a}{x}}$ $(n>3)$	0	$+\infty$	$\frac{a}{n}$	$\frac{a^2}{n^2(n-1)}$	$\frac{4a^3}{n^3(n-1)(n-2)}$
VIII	Skew	Un-limited	$k(a^2+x^2)^{-\left(\frac{n}{2}+1\right)}e^{\nu\tan^{-1}\frac{x}{a}}$ $(n>3)$	$-\infty$	$+\infty$	$\frac{\nu a}{n}$	$\frac{n^2+\nu^2}{n^2(n-1)}a^2$	$\frac{4\nu(n^2+\nu^2)}{n^3(n-1)(n-2)}a$

NOTES:—$\beta_1=\mu_3^2\div\mu_2^3$. $\beta_2=\mu_4\div\mu_2^2$.

Skewness $=$ (Mean—mode) $\div\sigma$

Criterion $=\mathrm{K}=\frac{\beta_1}{4(2-3\gamma)(4-3\gamma)}$

FREQUENCY CURVES.

($i.e.$, all >0), and in Types III, VI, VII and VIII, n must be >3].

μ_4	$\beta_1=\mu_3^2\div\mu_2^3$	β_2	$\gamma=\frac{\beta_1+4}{\beta_2+3}$	K
$\frac{3a^4}{(n+1)(n+3)}$	0	$3\left(\frac{n+1}{n+3}\right)$	$\frac{2}{3}\cdot\frac{n+3}{n+2}$ $>\frac{2}{3}$	0
$\frac{3}{4}a^4=3\mu_2^2$	0	3	$\frac{2}{3}$	0
$\frac{3a^4}{(n-1)(n-3)}$	0	$3\left(\frac{n-1}{n-3}\right)$	$\frac{2}{3}\cdot\frac{n-3}{n-2}$ $<\frac{2}{3}$	0
$\frac{48pq(2+\overline{n-6}pq)}{(n+1)(n+2)(n+3)}a^4$	$\frac{4(n+1)(p-q)^2}{(n+2)^2pq}$	$\frac{3[2+(n-6)pq](n+1)}{(n+2)(n+3)pq}$	$\frac{2}{3}\cdot\frac{n+3}{n+2}$ $>\frac{2}{3}$	$-\left(\frac{1}{4pq}-1\right)$ (negative)
$3m(m+2)a^4$	$\frac{4}{m}$	$3\left(\frac{m+2}{m}\right)$	$\frac{2}{3}$	∞
$\frac{48pq(2+\overline{n+6}pq)}{(n-1)(n-2)(n-3)}a^4$	$\frac{4(n-1)(p+q)^2}{(n-2)^2pq}$	$\frac{3[2+(n+6)pq](n-1)}{(n-2)(n-3)pq}$	$\frac{2}{3}\cdot\frac{n-3}{n-2}$ $<\frac{2}{3}$	$\frac{1}{4pq}+1$ (positive)
Not required	$\frac{16(n-1)}{(n-2)^2}$	...	$\frac{2}{3}\cdot\frac{n-3}{n-2}$ $<\frac{2}{3}$	1
$\frac{3(n^2+\nu^2)[(n+6)(n^2+\nu^2)-8n^2]}{n^4(n-1)(n-2)(n-3)}a^4$	$\frac{16(n-1)}{(n-2)^2}\cdot\frac{\nu^2}{n^2+\nu^2}$	$\frac{3(n-1)[(n+6)(n^2+\nu^2)-8n^2]}{(n-2)(n-3)(n^2+\nu^2)}$	$\frac{2}{3}\cdot\frac{n-3}{n-2}$ $<\frac{2}{3}$	$\frac{\nu^2}{n^2+\nu^2}$ $>0, <1$

Standard Deviation $=\sqrt{\mu_2}=\sigma$.

$$=\frac{\sqrt{\beta_1}(\beta_2+3)}{2(5\beta_2-6\beta_1-9)}.$$

$$=\frac{\beta_1(\beta_2+3)^2}{4(2\beta_2-3\beta_1-6)(4\beta_2-3\beta_1)}.$$